Macroeconomic policy and South African unemployment: developing a three-segment macroeconomic model

Philippe Burger and Frederick Fourie

Abstract

A recent survey of South African unemployment research reveals limited macroeconomic research on unemployment; in addition, almost all macroeconomic work on unemployment policy deals with the formal sector only. We critically evaluate the ability of standard macroeconomic theories to explain the South African unemployment situation and find that these theories provide a partial explanation at most. Even the few models that appear to incorporate an informal, or secondary sector, cannot explain persistent high unemployment or analyse labour flows between sectors/segments. To fill this gap we propose a macroeconomic framework that incorporates both formal (primary) and informal (secondary) sectors, imperfect competition, labour-union wage bargaining and labour-market entry barriers. We believe such a model provides a more suitable basis for macroeconomic policy analysis. The model shows, first, how a primary sector characterised by efficiency wage behaviour, a mark-up (due to high transport costs) and labour-union behaviour can explain the dual nature of the labour market – i.e. the existence of a secondary sector. Secondly, when barriers to entry prevent potential workers from entering the secondary sector, they end up being (involuntarily) unemployed in long-run equilibrium. The model potentially provides novel insights into labour flows between segments. Disturbances and fluctuations in the primary sector, for example, would spill over into the secondary sector and the third segment (comprising the unemployed). The paper concludes by suggesting a few policy implications and also a new macroeconomic research agenda on unemployment.

JEL codes: E24, E26, J2, J3

The **Research Project on Employment, Income Distribution and Inclusive Growth** is based at SALDRU at the University of Cape Town and supported by the National Treasury. Views expressed in REDI3x3 Working Papers are those of the authors and are not to be attributed to any of these institutions.

Macroeconomic policy and South African unemployment: developing a three-segment macroeconomic model

Philippe Burger and Frederick Fourie *

1. Introduction

1.1 The state of macroeconomic models and work on unemployment in South Africa

A recent survey of the South African unemployment research (Fourie 2011) reveals, inter alia, that there has been a limited amount of macroeconomic research on SA unemployment (compared to labour market studies, for instance). The existing work focuses on output elasticities of employment, sectoral shifts in employment, and possible labour-market related constraints on growth – and some work on increases in the long-run equilibrium rate of employment (the so-called NAIRU) after 1994.

As indicated by Fourie (2011), the South African policy debate on unemployment is encumbered by a divide that exists between macroeconomic and labour-economic analyses, in particular with regard to labour market segmentation and the role of the informal sector. Accordingly, a characteristic of all the macroeconomic work on unemployment in SA – and of macroeconomic policy analysis – is that it deals with the formal sector only. Meanwhile, evidence from unemployment research in the fields of labour economics and development indicate substantial segmentation in the South African economy: between the formal and the informal economies, within the informal sector, and between the unemployed and the informal and formal economies. Moreover, several labour market barriers exist that prevent people from improving their employment and earnings situation. Many such barriers impact especially the poor, and flow from the condition of poverty. This forms the basis of much marginalisation, inequality and continued poverty (Fourie 2011:10-44).

The labour market assumptions of standard macroeconomic models (and derived policy proposals) are at odds with such segmentation and barriers. The objective of this paper is to start bridging the divide between the macroeconomic discourse/models and the labour and development discourses on unemployment.[1] Presenting a concise critical evaluation of the

* Philippe Burger is professor and head of the Economics Department at the UFS. Frederick Fourie is Professor and Research fellow in the same department and Research Affiliate, SALDRU, University of Cape Town. Very useful comments from a REDI reviewer are gratefully acknowledged.

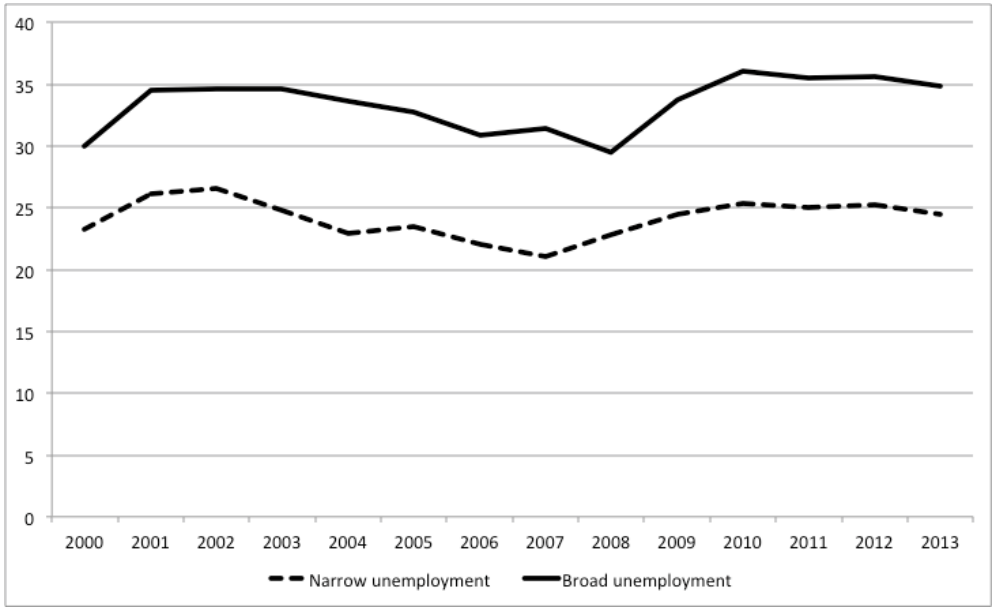
1 Some literature does exist in the field of development macroeconomics. Agénor and Montiel (1999) present a model incorporating a formal and informal sector. Basically it represents a model with traded and non-traded goods, with the former constituting the formal sector and the later the informal sector. The use of this model in South Africa is limited, as the informal sector, largely being retail-based, is a sector of traded goods.

ability of standard macroeconomic theories to capture and explain the South African unemployment situation and provide a basis for appropriate policy, the paper finds that at best these theories provide only a partial explanation. It proceeds to take up the challenge of including the realities of segmentation, dualism and labour-mobility barriers into a theoretical macroeconomic model. Initially this is done conceptually and diagrammatically and then in terms of a formal mathematical model. A major result of this model is that, given the incorporated features, it explains persistent high involuntary unemployment in equilibrium.

1.2 Compelling empirical dimensions of employment and unemployment in South Africa

Few countries have as serious an unemployment problem as South Africa. Graph 1 presents both the official (narrow) and broad unemployment rates for the period 2000-2013. In the period 2000-2013 the official unemployment rate never fell below 21% and averaged 24.1%. This 'narrow' rate only includes unemployed workers who actively search for a job. If the discouraged, i.e. non-searching, unemployed are also included, it shows that the 'broad' unemployment rate never fell below 30% and averaged 33.4% over the period.

Graph 1. Narrow and broad unemployment rates in South Africa 2000-2013



Data for 2000-2007 from the Labour Force Survey. Data for 2008-2013 from the Quarterly Labour Force Survey. All data refer to September of the relevant year. Source: StatsSA (2009, 2014)

According to Posel, Casale and Vermaak (2014) the search status of the unemployed does not predict their subsequent employment states, also meaning that the discouraged workers should not be excluded when measuring unemployment. Including discouraged workers indicates that the unemployment problem is significantly worse than suggested by the official definition.

Table 1 shows the composition of employment in South Africa, and in particular the existence of informal sector employment.

Table 1. Composition of the employed (% of total employment) 2000-2013

	Formal sector (Non-agricultural)	Informal sector (Non-agricultural)	Agriculture	Private households
2000	58.8	19.7	11.0	10.5
2001	63.5	18.3	7.4	10.9
2002	63.4	17.1	9.0	10.5
2003	65.3	16.8	7.5	10.4
2004	66.5	16.7	6.3	10.5
2005	65.0	19.8	5.4	9.8
2006	65.2	18.9	6.1	9.8
2007	68.7	16.3	5.7	9.3
2008	69.4	15.6	5.5	9.4
2009	70.8	15.2	4.9	9.1
2010	69.5	16.7	4.9	8.9
2011	70.8	16.0	4.6	8.5
2012	70.8	16.0	4.8	8.4
2013	71.2	15.4	4.9	8.4

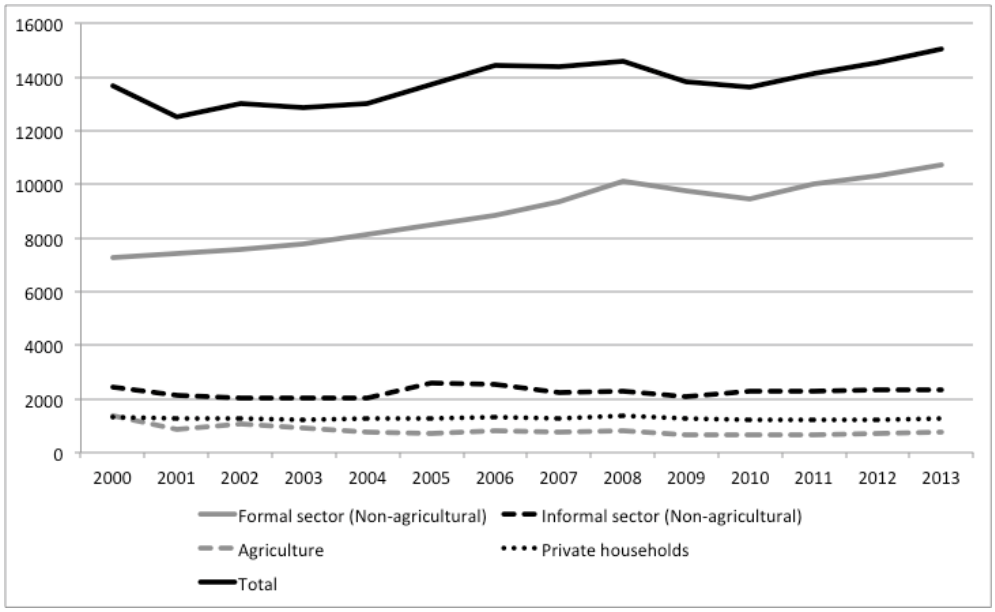
Data for 2000-2007 from the Labour Force Survey. Data for 2008-2013 from the Quarterly Labour Force Survey. All data refer to September of the relevant year. Source: StatsSA (2009, 2014)

A well-known peculiarity of South Africa is that, compared to peer-group countries, the informal sector is small relative to total employment (Kingdon and Knight, 2004). Moreover, it appears to have been shrinking in the past decade from almost 20% in 2000 to roughly 15% in 2013. Table 1 shows the falling share of the informal sector in total employment and the rising share of formal sector non-agricultural employment from roughly 59% in 2000 to 71% in 2013. Graph 2 shows that what happened is not that workers substituted formal for informal jobs: the number of workers employed in the informal sector remained more or less stationary while the number of workers in the formal sector increased.

In addition, and depending on which unemployment definition is used, there have been between two-and-a-half and three-and-a-half times as many unemployed people as informal sector workers (graph 2). Therefore, there is significant open unemployment.

This raises the following question: if workers do not find employment in the formal sector, why do they become unemployed rather than enter the informal sector? Kingdon and Knight (2004) suggest that there are significant barriers to entry into the informal sector, possibly in the form of capital and skills shortages. South Africa is not the only developing country where barriers to entry into the informal sector appear to exist. Grimm, Krüger and Lay (2011) and Grimm, Van der Hoeven and Ley (2011) find significant barriers to entry into the informal sector of many West African countries as well as Madagascar.

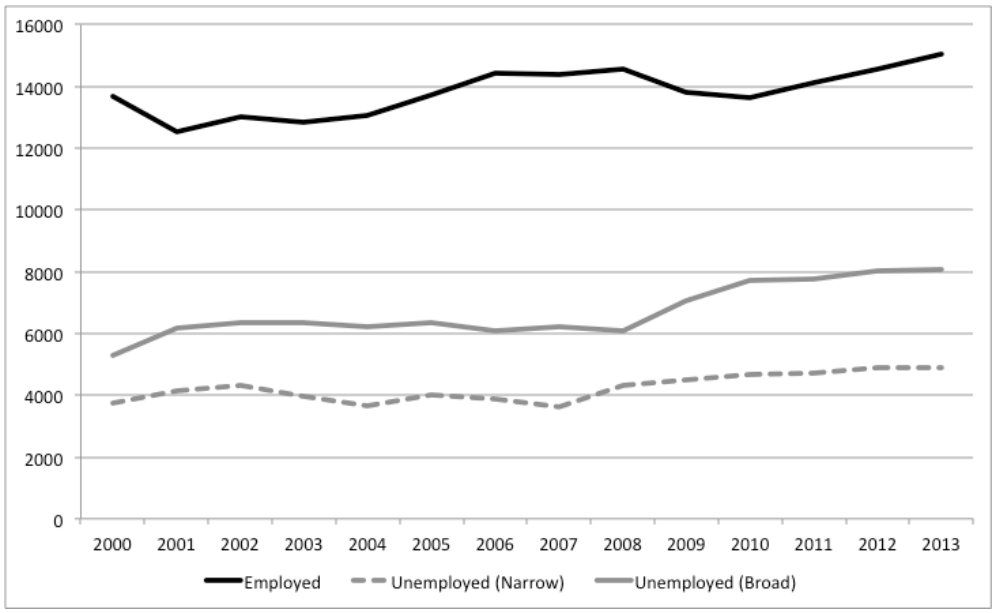
Graph 2 – The number of employed workers, formal and informal sectors



Data for 2000-2007 from the Labour Force Survey. Data for 2008-2013 from the Quarterly Labour Force Survey. All data refer to September of the relevant year. Source: StatsSA (2009, 2014)

The combination of an overall increase in the number of people employed in the formal sector since 2000 and an unchanging number of people employed in the informal sector would constitute a desired development were it not that the number of unemployed people also increased during the recession that started in 2007-8. This is shown in graph 3. (The increasing number of unemployed also increased the *rate* of unemployment, as shown in graph 1.)

Graph 3 – The number of employed and unemployed workers

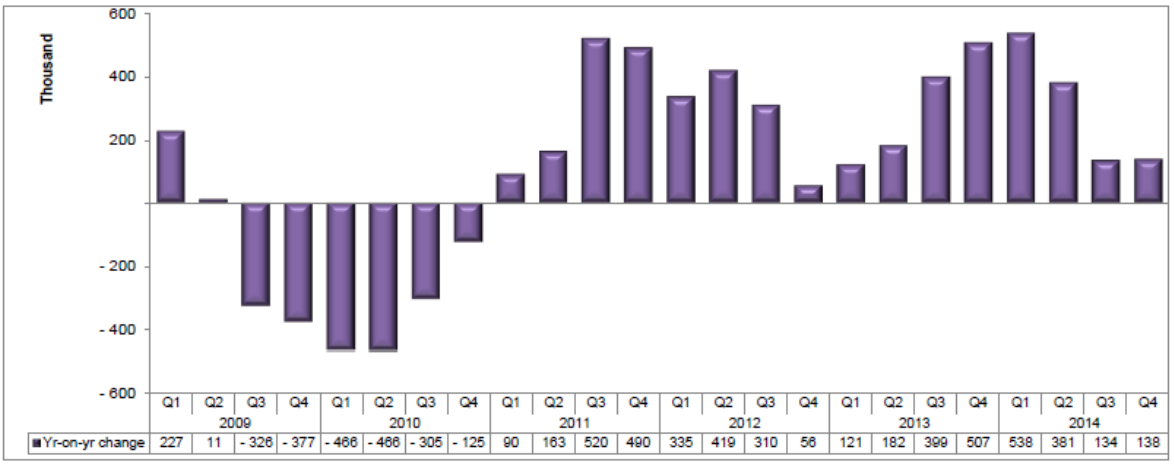


Data for 2000-2007 from the Labour Force Survey. Data for 2008-2013 from the Quarterly Labour Force Survey. All data refer to September of the relevant year. Source: StatsSA (2009, 2014)

What is noticeable is that, since 2007 the informal sector has not absorbed the additional people in the labour force who did not find employment in the formal sector. Informal sector employment numbers have not even increased proportionally to the growth in the labour force.

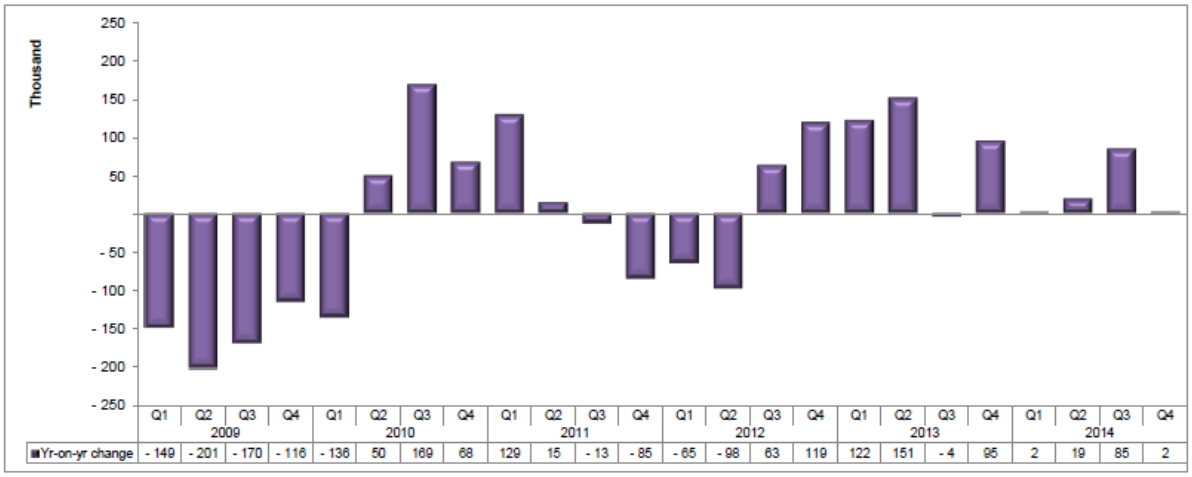
This suggests that, apart from longer-term trends, cyclical changes in formal and informal employment also are relevant, especially from a macroeconomic perspective. Graphs 4 and 5 show these:

Graph 4. Cyclical changes in formal sector employment (year-on-year)



Source: QLFS, 4th quarter 2014, Statistics South Africa

Graph 5. Cyclical changes in informal sector employment (year-on-year)



Source: QLFS, 4th quarter 2014,Statistics South Africa

It is apparent that, like the formal sector, informal sector employment also displays, and experiences, cyclical movements. Whether, and how, these may be linked, is outside the reach of current macroeconomic models and macroeconomic research. It is high time that models are developed that can start to address this gap.

Given the above numbers as background, a macroeconomic model that attempts to explain unemployment in South Africa needs to take account of the segmentation of the labour market. Such a model should further explain the persistent high unemployment rate and, more specifically, answer the question: if workers are not accommodated in the formal sector, why do they tend not to enter the informal sector, but rather remain unemployed?

1.3 Structure of the paper

Section 2 summarises and evaluates the labour component and unemployment analysis of mainstream theoretical macroeconomic models, in particular the currently dominant main alternatives: the New Classical model and the New Keynesian model (including multiple-equilibria and hysteresis models). Section 3 evaluates the applicability of these models in the South African, and possibly also developing country, context. The issue of segmentation appears to be critical, which leads to an evaluation of an often-quoted segmented macroeconomic model that has been proposed by Layard, Nickell and Jackman (LNJ). The LNJ model is not a fully-developed model, but rather a simple conceptual framework in diagrammatical format. Section 4 takes stock with some thoughts on the way forward. Section 5 then develops a conceptual and diagrammatical three-segment model – an expanded LNJ-model of sorts – that illustrates how labour market segmentation and labour market entry barriers can be integrated into the LNJ model. In spite of its diagrammatical nature, this expanded LNJ-model provides novel findings on labour market equilibrium in a segmented context and on the presence of voluntary and involuntary unemployment in such an equilibrium; it could also be used to show how supply and demand shocks differentially impact, through knock-on and flow effects, employment and unemployment in the three segments. The impact of hysteresis in such a model is also considered.

Against the conceptual background of section 5, section 6 develops a formal mathematical three-segment model in some detail. The presence of imperfect competition in product and labour markets, including the presence of labour unions, is a central feature. It builds on the seminal work by Bulow & Summers (1986) but also incorporates the suggestion of Kingdon and Knight (2004) and Grimm, Krüger and Lay (2011) that workers end up being unemployed and not in informal sector employment because of barriers to entry into the informal sector. As will be shown, the presence of al these elements implies the existence of a third segment that comprises the unemployed. In addition, in the long-run equilibrium there is involuntary unemployment.

On the basis of the potential of such an expanded model, Section 7 concludes by providing suggestions regarding priorities for macroeconomic research on unemployment and, by extension, poverty and inequality.

2. The labour market component of mainstream theoretical macroeconomic models

Like the traditional Classical/Keynesian divide of the 1930s to 1950s that turned into the Monetarist/Neo-Keynesian divide of the 1960s and 1970s, modern macroeconomics is deeply divided on the existence of involuntary unemployment. The modern debate is characterised by the divide between the New Classical and the New Keynesian models. The New Classical model has succeeded in making itself the reference model, putting the onus on others to argue and substantiate deviations from it.[2] Hence, it is also called the New Classical Benchmark Model (NCBM). This model posits, at its core, an economy in which there is no involuntary unemployment.

The New Keynesian model (NKM) finds itself on the other side of the divide, arguing for the existence of involuntary unemployment. It introduces imperfect product and labour markets as the key explanation for involuntary unemployment. More specifically, it allows for price- and wage-setting by firms and unions. Nevertheless, the NKM attempts to deviate as little as possible from New Classical model. Indeed, in several respects the New Keynesian model is but a variation of the NCBM. Pertinently, these models share a common analytical methodology for deriving macroeconomic behaviour from primary behaviour elements, i.e. the assumption of fully rational and optimising households and firms, linked to a labour market and a production function (see below).

The NCBM was also developed further into the Real Business Cycle Model (RBCM). The RBCM does not allow of involuntary unemployment. Thus, it finds itself on the same side of the divide as the NCBM.

The NCBM, the RBCM as well as the NKM are single-equilibrium models, i.e. models where output returns to its initial level after a demand shock causes it output to deviate from its pre-shock level. In the case of the NCBM and RBCM equilibrium occurs at full employment, with no involuntary employment, while in the NKM the long-run employment level allows for, and incorporates, involuntary unemployment.

Some extensions of the NKM are multiple-equilibrium models, i.e. models where, after a demand shock causes employment and output to deviate from their pre-shock level, they do not return to any particular level. A prominent example is the class of models characterised by hysteresis, which explains a long-run equilibrium with persistent involuntary unemployment.

[2] This can be compared to the situation in an earlier era when the debate was between Monetarist and Neo-Keynesian models. Keynes showed the inadequacies of the Classical model, and his model became the dominant (or reference) model. The Monetarists had to argue deviations from, or limitations of, Keynesian/Neo-Keynesian models.

It is clear that macroeconomists are deeply divided on the presence of involuntary unemployment. Together the NCBM and the RBCM constitute the view that unemployment is never involuntary. Economists who have adopted the NCBM and RBCM also view unemployment as much less of a problem than those who use the NKM. Those who accept the possibility of multiple equilibria see unemployment as even more of a problem.

However, as noted above, those models that incorporate involuntary unemployment equilibria define their models in terms of deviations from the NCBM. Therefore, the discussion of the range of models starts with a brief discussion of the NCBM, focusing only on its derivation and basic features.

2.1 New Classical Benchmark Model (NCBM) (The rational expectations model)

2.1.1 Deriving the NCBM

The NCBM is derived in a few simple steps (see Carlin and Soskice 2006: chapter 15). First a utility function and budget constraint is specified. Then the budget constraint is substituted into the utility function, which is then solved with respect to the employment variable. This produces the *labour supply* function. (Alternatively, setting wages in the labour supply function as left-hand variable for periods t and t+1 allows for the derivation of the *Phillips curve*.) Using the production function to derive the marginal product of labour yields the *labour demand* function. The labour supply and demand functions together yield the equilibrium level of employment and wages.

2.1.2 Basic features of the NCBM

In the NCBM microfoundations are explicit. The model assumes rational expectations (i.e. all agents have full information on the 'true' model of the economy) and there is a total absence of any nominal rigidities of wages and prices (no contracts). Thus, the model implies an immediate adjustment to equilibrate supply and demand in markets.

The labour market component of the model is based on a labour supply (L^S) and labour demand (L^D) component where labour demand is determined by the marginal product of labour (Carlin and Soskice 2005: 568; 580). Long-run equilibrium in the model is supply-side determined, with a rapid return to long-run equilibrium (or NAIRU) after a disturbance. Thus, there is no distinction between short-run and long-run equilibria.

In equilibrium there is no involuntary unemployment – only frictional unemployment – and therefore no unemployment problem: "No unemployed persons in this model, only non-participants" (Cahuc and Zylberberg 2004:459).

Aggregate demand (and demand policy) can have no effect on equilibrium employment, not even in the short run. However, as pointed out by Cahuc and Zylberberg (2004:455): "This

prediction is not verified by data". A Phillips curve exists in the model, but due to rational expectations the relationship cannot be exploited systematically. Only unexpected changes in demand can lead to temporary Phillips curve effects (i.e. using demand policy to reduce unemployment in return for accepting higher inflation).

2.2 Real business cycle model (RBCM)

Deriving the RBC model is straightforward and based on the NCBM (see Carlin and Soskice 2006: chapter 15). It is the intertemporal re-specification of the classical model and therefore represents an extension of the NCBM. As such, it shares the same microfoundations with the NCBM.

A distinguishing feature of the RBCM is that business cycles mainly are modelled as fluctuations *of* the equilibrium (rather than deviation from equilibrium), meaning that business cycles are seen and modelled as equilibrium phenomena. It also means that another feature of RBCM is that of continual optimality. Equilibrium employment and output vary over time due to technological and other supply-side shocks. These variations in the equilibrium constitute the business cycle. In equilibrium (which virtually always exists) there is no involuntary unemployment – only frictional and search unemployment – and whatever unemployment level transpires, is optimal, since it reflects the optimising responses of agents to technology or preferences. Therefore, strictly speaking all unemployment is voluntary. Since all unemployment is voluntary, in this theoretical model there is no unemployment problem, no business cycle problem and hence no stabilisation role for macroeconomic policy.

2.3 New Keynesian model (NKM)

2.3.1 Deriving the NKM

As mentioned above, the NKM is defined in terms of its deviation from the NCBM (see Carlin and Soskice 2006: chapter 15). It is derived using the same steps as for deriving the NCBM. However, it introduces imperfect competition (monopolistic competition) through price-setting behaviour of firms, which is substituted into the constrained utility function as used in the NCBM. (It enters through the equation "wage = price minus mark-up", where the mark-up depends on the elasticity of demand – i.e. a typical monopoly model). Thus, the New Keynesian Model is derived by amending the NCBM – it is a "spanner in the works" model.

2.3.2 Basic features explaining long-run involuntary unemployment

The NKM also has explicit microfoundations. Like the NCBM and the RBCM, it assumes *rational expectations*, with rational intertemporal optimising behaviour displayed by all agents. However, the model differs from the NCBM and RBCM by accepting price setting

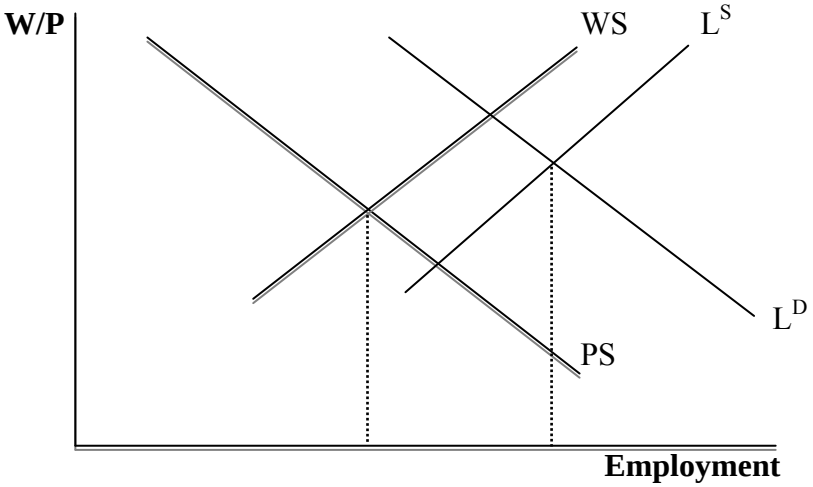
and wage setting in *imperfectly* competitive markets. This implies rigid or slowly-adjusting prices and wages. Nevertheless, it can be said that the NKM wholly adopts major elements of the New Classical model and that there is not much 'Keynesian' about it (also with regard to aggregate demand effects; see below).

The New Keynesian model strives to deviate as little as possible from the New Classical model, but does insist on inserting the imperfect competition assumption. Using this assumption the model derives its important involuntary unemployment result.

As in NCBM the NKM long-run equilibrium employment level (NAIRU) is supply-side determined. However, assuming imperfectly competitive goods and labour market conditions (and thus inflexible prices and wages), the model results in a long-run equilibrium with involuntary unemployment (in addition to frictional and search unemployment). Therefore, persistent unemployment exists even in the face of rational expectations and maximising agents. Indeed, it results from such behaviour.

Two types of models can be differentiated on the basis of the imperfect market conditions included. In the first type firms are price setters while unions are wage setters. In the second type firms are both price *and* wage setters. Such price- and wage-setting behaviours change the model significantly. Instead of labour supply and demand, the NKM uses wage-setting (WS) and price-setting (PS) relationships to derive labour market equilibrium. Diagrammatically WS is upward-sloping, while the PS downward-sloping.[3] Figure 1 demonstrates that the WS-curve lies above the Classical L^S-curve, and the PS-curve lies below the L^D-curve.

Figure 1: The NKM labour market and the PS–WS system



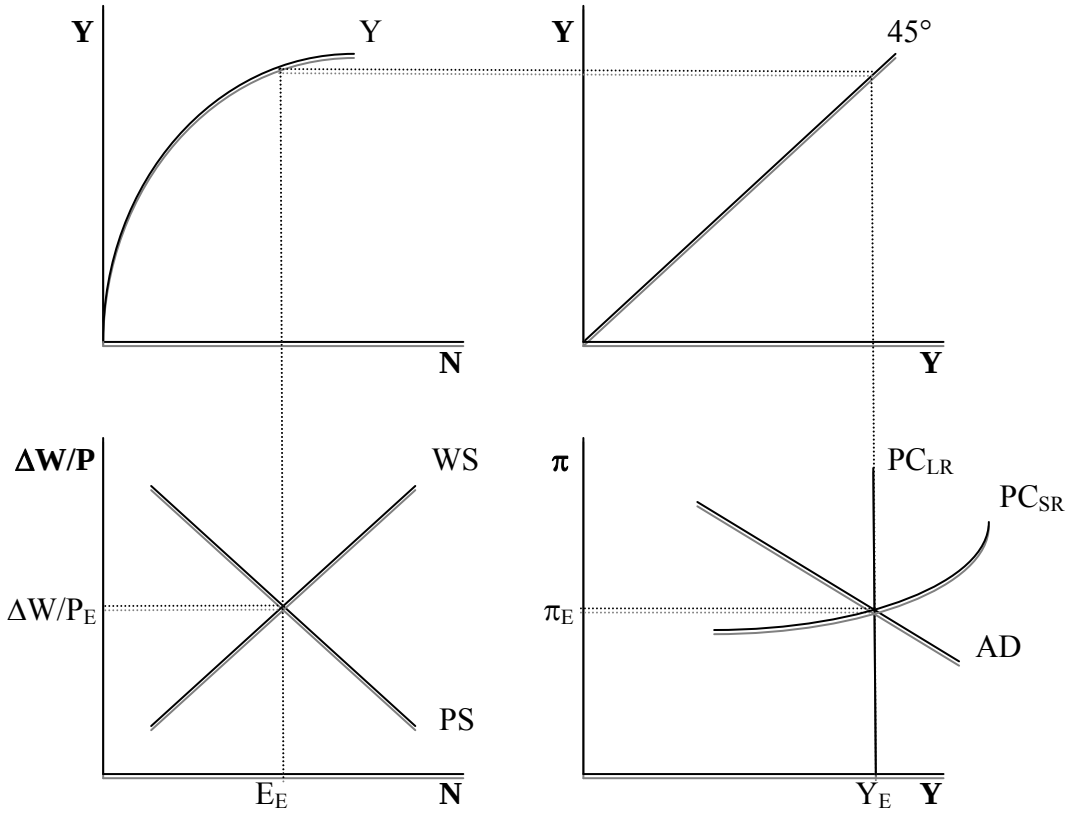
[3] That is if diminishing marginal returns to labour (MPL) exists; if MPL (and the mark-up) is constant, PS will be horizontal (Carlin and Soskice 2005: 49; Cahuc and Zylberberg 2004:381).

The long-run equilibrium level of employment is below the competitive/Classical equilibrium level. The difference between the two equilibrium employment levels represents long-run involuntary unemployment, caused by imperfect competition conditions.

To complete the model, the labour market and resulting employment level are linked to output via a production function, as demonstrated in figure 2.

- Note that, unlike figure 1, where the real wage appears on the vertical axis of the WS-PS system, the WS-PS system in figure 2 has the *rate of change* in the real wage rate on the vertical axis. This allows putting the inflation rate instead of the price level, in relation to output, in the bottom right-hand graph. With inflation on the vertical axis the AS curve becomes the Phillips Curve.
- Figure 2 also presents a short-run Phillips Curve that has a positive slope, which gets steeper at higher output levels. The steepness results from the assumed decreasing marginal product of labour that also causes the production function in the top-left graph to increase at a decreasing rate, and the PS in the bottom-left graph to have a negative slope.

Figure 2: PS and WS, output and inflation



In the NKM, profit-maximising wage-setting and price-setting behaviour determines the equilibrium wage rate and the equilibrium employment level E_E. Because of factors such as

the menu costs of price setting, bargaining costs of wage setting, and the use of long-term, multi-period contracts, wages and prices adjust slowly (called wage and price stickiness). Therefore persistent involuntary unemployment is not a pure labour market problem caused by excessive wages or, for example, labour market regulations. It is the outcome of rational optimising behaviour of firms and consumers/workers in a specific competitive context (i.e. imperfect competition).

Contrary to earlier Keynesian and Neo-Keynesian models, aggregate demand deficiency cannot be the cause of a long-run equilibrium with persistent/long-run involuntary unemployment in the NKM. Should a demand deficiency occur, the short-run Phillips curve will adjust (as expectations catch up with actual inflation) and return the equilibrium to the vertical, long-run Phillips-curve and long-run income equilibrium at Y_E. In this respect the NPK fully accepts the NCBM result on aggregate demand impotency in the long run.

2.3.3 The short run: business cycles and the Phillips curve

The NKM differs significantly from the former two models in terms of its treatment of the short run. In the NKM the business cycle is explained as fluctuations *around* the equilibrium. Business cycles in the NKM are interpreted as *dis*equilibrium phenomena (i.e. disequilibrium relative to the long-run equilibrium and WS-PS intersection) – in contrast to the RBCM where business cycles are seen as movements *of* the long-run equilibrium employment and output levels themselves (and therefore are seen as equilibrium phenomena).

In the NKM short-run or cyclical unemployment is due to demand or supply shocks; combined with expectational mismatches due to existing contracts. In addition, imperfect competition means that quantity (i.e. output and employment), rather than prices and wages, changes first (after a demand shock, say). Thus, adverse demand shocks reduce output and employment in the short run. The effects of these shocks might be either amplified through multiplier effects that operate through consumption, or dampened because of anticipatory consumption smoothing. There is a slow return to long-run equilibrium / NAIRU (due to, inter alia, inflation inertia (Carlin and Soskice 2005:595)). This slow return creates room for a short-run Phillips curve, with the NKM tradition containing various forms of the Phillips curve (including sticky prices with staggered price setting and sticky information resulting from delayed information on monetary policy).

2.3.4 Two subtypes of NKM models

As noted above, in the NKM tradition two types of models can be differentiated on the basis of the imperfectly competitive market behaviour that they include.

a. The first type is a class of models with firms as price setters and unions as wage setters (‘union models’). In the union model monopolistic firms use mark-up pricing while the union acts as a monopolistic wage setter.

b. The second type is a class of models with firms as both price and wage setters (efficiency wage models). In the efficiency wage models monopolistic firms use mark-up pricing *and* set efficiency wages.

A mix of these model types can also be conceived, as will be demonstrated below.

A key feature of the efficiency wage models is the introduction of the assumption that worker productivity is a function of wage levels. At each level of employment firms set the wage above the wage at which workers are willing to supply their labour. In graphical terms, the wage offered by the firm at each level of employment is above the L^S curve and thus above any L^S-L^D equilibrium wage. Firms offer these higher wages because higher wages increase the potential loss that workers could incur if caught shirking. Hence, higher wages serve as an incentive to be more productive. Generally the focus is on the upper range of wages and skills, where the wage-productivity nexus is assumed to be important.

2.4 Multiple equilibria models

2.4.1 Basic features

All the above models are single equilibrium models in the sense that those factors that might cause the WS and PS functions to shift (such as the mark-up by firms, the level of union power or the unemployment benefits paid by government), are assumed to be variable only in the long run. Thus, in the above models, long-run equilibrium (un)employment can shift over time due to various types of structural or institutional changes that impact WS and PS (creating a "time-variant NAIRU").

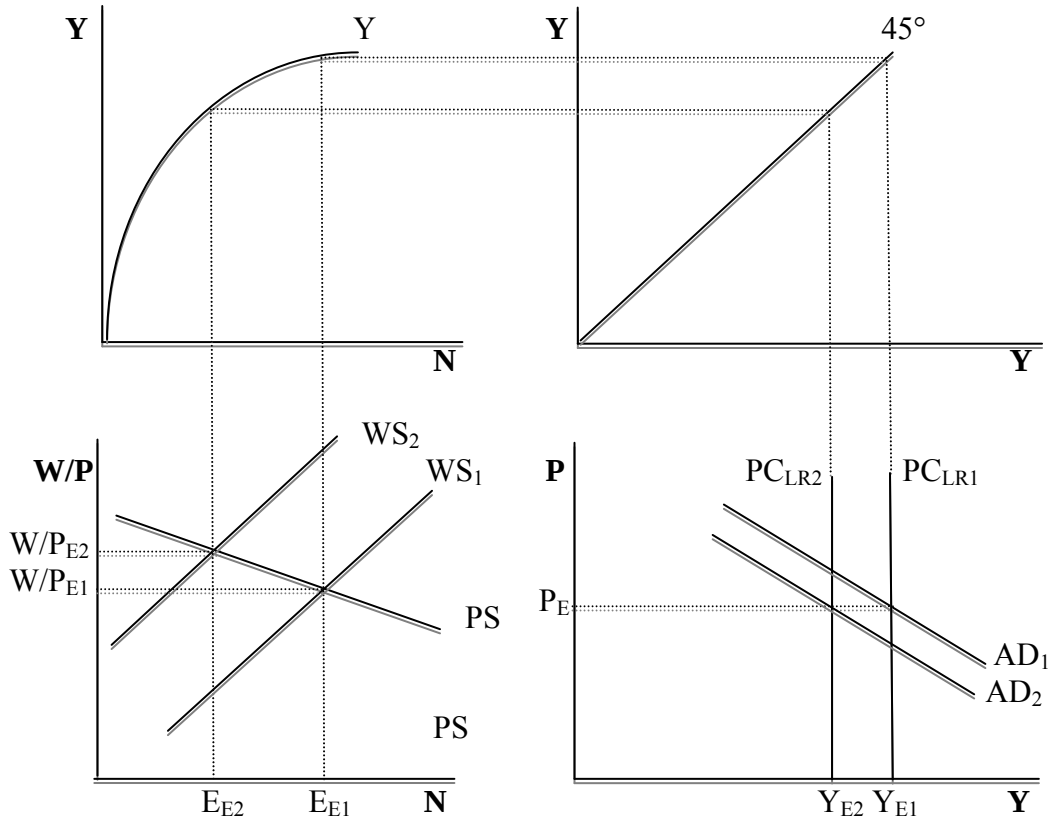
Multiple equilibria models add a further dimension to structural or institutional changes that may impact on wage- and price-setting behaviour. However, the structural changes added by the multiple equilibrium models are assumed to have a higher frequency of occurrence than the structural or institutional factors noted above. It is this higher frequency that leads to these models being denoted as 'multiple' equilibrium models – the changes occur in what might be denoted the medium term. The best known multiple equilibria models are hysteresis models. These models were mainly developed to help explain persistent unemployment (especially in the European context; see Blanchard (2005) and its list of references). Other models are based on pro-cyclical labour productivity ("high-road/low-road" models) or on the role of fairness in the utility function of workers.[4]

[4] The "high-road/low-road" models analyse the existence of multiple equilibria on the basis of pro-cyclical labour productivity. At high levels of income, productivity is high, and so are wages and employment, while at low levels of income, productivity is low, and so are wages and employment. (Exogenous shocks and an efficiency wage argument could explain why an economy ends up in the high (low) income, productivity and employment range.) In this model the economy can get stuck in a low-employment equilibrium (or a high-employment equilibrium). Another type of multiple-equilibria model postulates the existence of a range of equilibrium (un)employment rates that each depends on the role of fairness in the utility function of workers.

2.4.2 Hysteresis models

The hysteresis model is an extension of the NKM model (see Carlin and Soskice 2006: chapter 15). The hysteresis model incorporates the monopolistic price setting behaviour of firms and typically also the wage setting behaviour of unions. As is the case with the standard New Keynesian model, the hysteresis model allows for involuntary unemployment. In addition, it explicitly considers multiple equilibria by focusing on how the economy can move from one equilibrium to another. Once an equilibrium employment level is disturbed by a demand shock, employment can settle at a new equilibrium level and therefore display no tendency to return to its previous equilibrium level. Higher actual unemployment determines a new equilibrium unemployment level – therefore, the equilibrium is a function of the history of unemployment. This implies unemployment persistence. Moreover, not only does the hysteresis model explain the persistence of high unemployment and the presence of involuntary unemployment, it also implies that involuntary unemployment can occur at any level of unemployment.

Figure 4 – Hysteresis due to skills loss



There are various versions of the hysteresis model (Carlin and Soskice 2005: 117-20; 617-20). The first version is based on deskilling. Unemployment is initiated by a fall in aggregate and hence labour demand, but perpetuated by the loss of skills and productivity caused by the higher actual unemployment. The deskilled unemployed become less employable, which

translates into less competition for jobs. Graphically, the WS shifts left and up (figure 4, bottom left quadrant). Equilibrium employment decreases from E_{E1} to E_{E2} and long-term equilibrium unemployment increases. There is no convergence or return to a stationary equilibrium (Cahuc and Zylberberg 2004: 480). When demand is restored, the unemployed do not necessarily find employment again. The equilibrium is at a lower level of employment.

Note that this process is asymmetric. While a negative demand shock might cause the supply of labour to decrease (because skills decrease), a positive demand shock will not cause the supply of labour to increase. Because of a lack of skills the supply of labour is constrained, meaning that, following the positive demand shock, there is a shortage of skilled workers to employ. This might thus explain why a restoration of demand after a slump does not necessarily translate into more employment.

The implication of this model is that changes in the wage depend not on the stock of unemployed, but only on changes in the number of short-term unemployed (Cahuc and Zylberberg 2004: 480). Thus, the long-term unemployed do not affect wages, they are outside the labour market (Cahuc and Zylberberg 2004: 480) and they are involuntarily unemployed. Therefore, this is a form of segmentation, a case of the 'heterogeneity of the unemployed' In so far as one can still call it a *long-run* Phillips curve, an increase in the long-term unemployed will cause a fall in the long-run output level, which will cause the long-run Phillips curve to shift left (figure 4, bottom right) (it might now be more apt to call the vertical Phillips curve the medium to long-run Phillips curve).

A second version of the hysteresis model focuses on the scrapping of capital. The increase in the long-term unemployment is initiated by a fall in aggregate and hence labour demand, but is perpetuated following the scrapping of capital, itself a goods market supply-side phenomenon caused by the fall in aggregate demand. The lower level of capital limits the extent to which the goods market can react when aggregate demand returns to the level existing prior to the decrease in demand. As a result, employment does not return to the level existing prior to the decrease in demand. Graphically the PS curve shifts downwards (if it has a negative slope), or is truncated (if it is flat). As demand picks up again, profits rise, but no new capacity is added in the immediate aftermath of the rise in demand. In the long-run, though, it will be profitable to add new capacity. Thus, this is seen as a quasi-hysteresis in that long-run unemployment increases in reaction to an increase in actual unemployment, but when in future companies start adding new capacity, long-run unemployment will decrease.

The third version of the hysteresis model focuses on insider-outsider union behaviour, where insiders take active steps to exclude outsiders from the work place. Long-term unemployment is, as in the above models, initiated by a fall in aggregate demand and hence labour demand, but perpetuated by the insider-outsider behaviour within labour unions. This is graphically

represented by a kinked WS-curve that is positive up to the point of inflection, where after it becomes vertical.

Note that in all the hysteresis models an aggregate demand shock is the initiating factor, but the *persistence/perpetuation* of the higher unemployment (i.e. its hardening into a new, lower-employment equilibrium) is due to supply-side behaviour in either goods or labour markets. Such goods and labour market supply-side behaviour comprises either firms that scrap capital, workers who lose skills or unions that exclude the unemployed. The results imply that aggregate demand, by affecting the short-run equilibrium, can alter long-run equilibrium (un)employment. Thus, in hysteresis models transitory shocks may have permanent effects.

3. Applicability of the mainstream models to the South African situation

This section considers the applicability of the labour component of the above models to the South African and developing country situation.

3.1 NCB/RBC family

It is apparent that there are not many places in these models where characteristics peculiar to the South African labour market situation – noted in the Introduction – can be inserted or incorporated. In considering applicability, one must remember that in terms of the accepted methodology these models are not supposed to be realistic – they adopt a methodological simplification intent on demonstrating that undisturbed markets yield optimal outcomes in terms of maximised utility. In these models there is only one (i.e. formal) sector; there is no secondary or informal sector. Those who are unemployed are assumed to have *chosen* not to be employed at the going market wage – they are *voluntarily* unemployed. These unemployed simply are *not in the labour force.* In addition, those who cannot find employment have the option of turning to self-employment.[5]

The value of this model, which explains unemployment as a voluntary condition, seems rather limited in a country where according to the narrow definition 25% of the labour force is unemployed.

Nevertheless, the RBC model might provide interesting insights as it highlights the impact of technological and supply-side shocks. It could find potentially interesting applications in a developing country that is subject to, for example, technology import shocks of which many are labour-saving and high-skills demanding. However, in RBC models, just as in NCB

[5] Self-employment comprises being an owner-manager of a one-person or multi-person (i.e. employing) firm. (In labour-market literature the former is also called an 'own-account worker'.)

models, labour is mobile and wages adjust quickly, leaving only the voluntary unemployed after the adjustment.

3.2 NKM family

Models in this family have undergone much refinement to explain especially European unemployment and hysteresis ('Eurosclerosis' in the 1980s and 1990s and after 2008). These refinements reflect imperfect competitive behaviour (price and wage setting behaviour) and more realistic behavioural assumptions (e.g. the wage-productivity link in efficiency wage models).

Given the oft-cited high levels of concentration in many South African goods markets, the price-setting model (with mark-up behaviour by companies at its core) appears relevant to the South African situation. The union wage-setting model also appears relevant to the South African situation, especially variations on the insider-outsider model.

Efficiency wage models may also be relevant – South African firms may behave no different from firms elsewhere. However, given the existence of strong unions in many industries, it is a question whether firms are wage setters in South Africa – suggesting that the union model might be more applicable. (In addition there is also minimum wage-setting for non-unionised sectors via the Basic Conditions of Employment Act.) In mixed model with both efficiency wages and union wage bargaining provides an enticing option in the South African case – something we explore in our model below.

3.3 Segmentation? Weaknesses of the NKMs

Although the New Keynesian models appear to be more applicable to a situation such as that of South Africa, they have several limitations. First, all these models are only applicable to the primary (e.g. formal) sector. Nevertheless, there are hints at segmentation:

- Efficiency wage setting by firms implies a primary and a secondary sector.
- Union wage setting implies insider-outsider situations.
- Skills loss by long-term unemployed (heterogeneity of the unemployed) implies that wages are not influenced by the stock of such long-term unemployed. This suggests that these *long-term unemployed are outside the formal labour market*.

However, even though such elements imply a kind of segmentation, no attention is paid to the implications of segmentation for employment and unemployment. These models are silent on the secondary market or the 'outsider' market or the survival of the long-term (involuntary) unemployed. The economic activities of those in the secondary sector are treated as a residual not requiring further analysis or explanation. However, this raises the question whether secondary sector agents are not 'rational and optimising' like primary sector agents? If they

are rational and optimising (and why would they not be?), why not analyse the secondary sector explicitly?

The NKM and hysteresis models both can explain why workers lose employment, whether in the short or long run. However, what these models do not consider or explain is why those who lose employment then become unemployed and not self-employed. To do so, these models, will have to generalise the basic skills argument of the hysteresis model and state that those employed have a different skills set than those who are self-employed. Thus, once workers lose their job, their skills set (aimed at employment) might not allow them to become self-employed. In addition, the continued depreciation of their employment skills set means that after a while they have no relevant skills set left, and hence remain unemployed. Note that this weakness is not only a weakness in explaining unemployment in South Africa, but also in a European/First World context.

A failure to explain the imperfect substitution between employment and self-employment is a weakness in most of these models. Indeed, the proper and refined treatment of self-employment is a substantive gap in most models. Some models ignore it, while others summarily include it in the category of 'workers' (normally in a secondary sector) – ignoring the question, for example, whether self-employment is a labour market issue or an emerging firm issue, and what the dynamics between these two issues may be. This may be quite pertinent in a developing country context.

3.4 The segmented labour market model of Layard, Nickell and Jackman (LNJ)

This section discusses an explicit segmented market macro model. It is the model of Layard, Nickell and Jackman (1991:41-44; also 2005), the most well-known attempt to incorporate labour market segmentation into a macroeconomic model.

3.4.1 Basic features

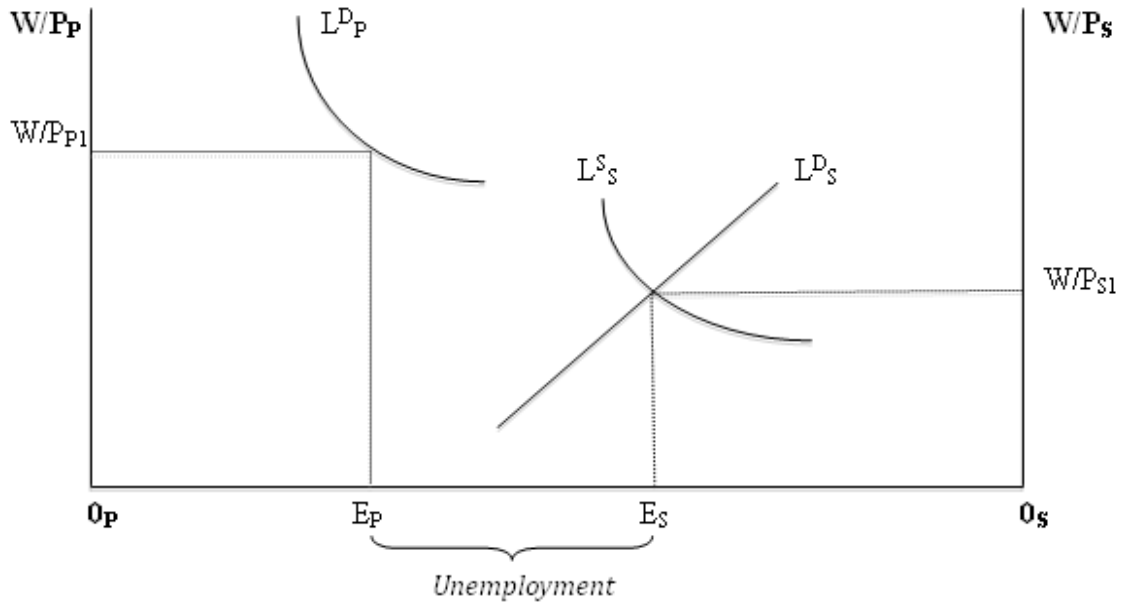
Segmented labour market models typically include a primary and a secondary sector. The LNJ model does the same. With respect to the primary sector its features are typically New Keynesian. They do not use the WS-PS framework, but their model can be translated into this framework without any harm (see section 5 below).

In the primary sector, wage setting is done by firms through either efficiency wages or via union bargaining mechanisms, or a combination of these - both unions and employers have incentives to set wages above market-clearing levels. In terms of figure 5, wages in the primary sector are set at W/P_{P1}. Note that because the wage is set above the labour market-clearing level, only points on L^D_P are relevant. Wage setting determines primary sector employment. Reducing primary sector wages would increase labour demand and employment. In equilibrium there are not many vacancies, and no shortage of labour for jobs. The

excess primary sector labour supply – those excluded (or rationed) from the primary sector – goes to the secondary sector. In figure 5 the secondary sector is measured on the horizontal axis from zero on the right-hand side. (The length of the horizontal axis indicates the entire labour force, i.e. the economically active.)

While the primary sector is New Keynesian in nature, the secondary sector is surprisingly very New Classical. The secondary sector labour market is assumed to be market clearing "in the sense that wages are not high enough to attract a queue of job-seekers, nor do vacancies last long since skill requirements are low" (Layard et al 1991: 42). In figure 5 the wage in the secondary sector adjusts to clear at W/P_{S1}, where labour supply in the secondary sector, L^S_S, equals labour demand L^D_S. Layard *et al* include self-employment in their definition of employment, so the voluntarily unemployed are not self-employed. If secondary sector wages were lower, employment would fall (due to a reduced supply of labour). The rising labour supply curve is generated by a range of reservation wages of different people in the secondary sector.

Figure 5 – The LNJ two-sector model



Equilibrium produces an employment level in each sector, presumably feeding into corresponding production functions and a combined aggregate output. Employment also produces a group of economically active people that are 'between the two sectors' – the distance between E_P and E_S – that are unemployed. In a sense they are *both voluntarily and involuntarily unemployed:* they are 'involuntarily unemployed with respect to primary sector' at the going wage there, but simultaneously they are 'voluntarily unemployed with respect to the secondary sector' (i.e. not willing to work at the going wage in the secondary sector). In the final instance they are voluntarily unemployed.

If shocks occur that affect primary sector employment, those shocks will be reflected in changes in secondary sector labour supply and employment as well as in unemployment. Thus there are spill-over effects from the primary sector. (Such dynamics will be analysed further in section 5 below.)

Compared to the standard New Keynesian models the LNJ model introduces the possibility that labour market conditions are not homogenous across the economy. More specifically, their model allows of the existence of a persistent wage differential in the economy, with wages in the primary sector being higher than wages in the secondary sector. The New Keynesian features of the primary sector cause a lower employment rate in the primary sector, with those who would have been employed in the primary sector now supplying their labour in the secondary sector. This additional labour supply in the secondary sector causes wages in that sector to be lower than those in the primary sector. (In the absence of the New Keynesian features, wages theoretically would be expected to be uniform across the two sectors, rendering the distinction between the two sectors redundant).

3.4.2 Applicability of the LNJ type of segmented model to SA and developing country situations

The LNJ model appears to be a promising improvement on the standard New Keynesian model. The model explicitly recognises segmentation and shows the existence of a secondary labour market, unlike all the other models. But there is limited analysis of the labour market behaviour of those in the secondary sector. And the LNJ model suffers from the same New Classical critique: Why do many of those who become unemployed in the primary sector, stay unemployed and not all become (self-)employed in the secondary sector?

Even though the LNJ theoretical model contains a sector with New Keynesian features, as in the one-sector New Keynesian models these features do not produce or explain the existence of involuntary unemployment at the aggregate level (though workers can be involuntary unemployed with respect to the primary sector). The assumption of a perfectly competitive, market-clearing secondary labour market delivers this result. In this sense the Layard *et al* model ends up being quite close to the NCBM in which there is also no aggregate involuntary unemployment. However, as will be demonstrated below (section 5), the unemployment rate in the model exceeds the unemployment rate that would exist in the absence of New Keynesian features.

In addition, while Layard *et al* introduce their much-quoted segmented-model diagram in the first, overview part of book, they immediately *abandon* it and proceed to in-depth analysis of the primary sector only. They are short-lived segmentationists. They offer no further analysis of labour market behaviour of those in the secondary sector. Also, there is no analysis of the flows of labour between the two sectors – which, potentially, is crucial to understanding

employment and unemployment. For example, what are the determinants and dynamics (and possible hindrances) of such transitions?

Moreover, there is no analytical mention, appreciation or consideration of the economic activities of those excluded even from the secondary market: the unemployed who find themselves in a *tertiary* segment (the 'default' segment) and, obviously, no analyses of their transitions to and from the secondary or primary sectors. They simply are seen as voluntarily unemployed and thus unproblematic.

4. Overall evaluation and way forward: two key issues

The 'standard' (mostly New Keynesian) macroeconomic theories used internationally to explain unemployment, such as the efficiency wage model or the labour union bargaining model, assume that product markets are not fully competitive. The resultant price-setting (as opposed to price-taking) behaviour of firms allows them to pay higher wages, usually a result of efficiency-wage or union bargaining behaviour. Since the formal sector in South Africa is characterised by a high degree of concentration (Fedderke and Naumann 2011; IMF 2013), one might expect these theories to explain some of the unemployment in South Africa.

However, from a macroeconomic point of view such standard theories only provide a partial explanation of the peculiar unemployment problem encountered in South Africa. In a standard model, workers who are not employed in sectors characterised by efficiency wage-setting behaviour would be expected to seek (and find) employment in sectors not characterised by such behaviour – in particular, in the informal sector, which is not characterised by high concentration or unions or by large firms and economic concentration (therefore, efficiency-wage and labour-union determined behaviour is not expected there). However, the data show that a large proportion of such workers are not accommodated in the informal sector and end up being unemployed. This is not explained by the standard theories.

The point is that, at present, macroeconomic theory from whichever school of thought analytically more or less ignores the existence of a secondary sector – and it has absolutely no theory of the tertiary segment (i.e. those who drop out of the first two segments altogether). Consequently there is no theory of inter-sector labour flows and their determinants and dynamics. In addition, there is also no theory to explain the imperfect substitutability between employment and self-employment that can help answer the New Classical critique, which is: why do workers who lose employment become unemployed and not self-employed? Is it simply a choice due to earnings being lower than reservation wages – or are there barriers and constraints that explain the situation?

Two key issues arise. The first relates to the need to explicitly account for segmentation. It is necessary to open the analytical box with regard to the *secondary* and *tertiary* segments, i.e.

to analyse and model economic and labour market behaviour within all segments. At the very least the secondary sector should be analysed in depth. There also is a need to analyse flows of labour between these sectors/segments – this is absolutely crucial to understanding employment and unemployment. What are the determinants and dynamics of such flows? Can they be modelled?

The second issue relates to models that can account for a non-clearing secondary labour market, and thus long-run involuntary unemployment. The latter implies the consideration of barriers to entry and mobility barriers between the sectors/segments. None of the models discussed above allow for the possibility that entry and mobility barriers might exist for those who wish to enter into employment in either the formal or the informal sector.

This is particular relevant for application to the South African context given that the existence of a range of entry and mobility barriers is central to the findings from much South African labour market and development literature on unemployment and employment (cf. Fourie 2011: 41-44). These include skills mismatches, geographical-spatial factors such as transport costs, lack of work experience, household culture with respect to work experience, work ethic and search; lack of information about jobs and jobs environment, lack of labour market networks, lack of resources to support search. Entry and mobility barriers mean that labour supply is not merely a function of wages. Any job search activity implies a balancing of risks/expectations and costs. Search barriers and costs raise the reservation wage for job searchers, restrain and, beyond some level, truncate labour supply.

Thus it is necessary to incorporate entry and mobility barriers that affect transitions between the three segments – notably from tertiary to secondary, and then from secondary to primary segments. In this way the analysis of barriers links up with the analysis of inter-sectoral flows. In the discussion below, and specifically in the exposition of the mathematical model, two specific types of barriers will receive specific attention given their relevance in the South African labour market setting. Both are financial, with the first relating to high transport cost and long traveling distances to places of work, and the second relating to a lack of funding to set up the (physical and human) capital needed to operate in the secondary sector.

The next section develops a conceptual and diagrammatical model – roughly along the lines of the LNJ model but substantially augmented in a way that starts to include some of these dimensions of the South African labour market.

5. A conceptual three-segment 'barrier' model: augmenting the model of Layard *et al.*

5.1 Outlines of the model structure

The basic conceptual framework developed here is a *three-segment model* comprising:

(a) a *primary sector*, defined as the employment (or self-employment) sector of choice for all workers (i.e. what could be termed a 'secure jobs' sector);
(b) a *secondary sector* that serves as the second-choice employment (and self-employment) sector for workers (i.e. a 'less secure jobs' sector), and
(c) a *tertiary segment* which contains the unemployed (who survives in various ways).[6]

The stylised characteristics of these segments can be described as follows:

(a) *Primary sector (as in New Keynesian one-sector & LNJ two-sector models):* This sector largely provides work stability and security for those who are employed and self-employed. The sector comprises firms that act as monopolistic competitors and therefore price setters; wage setting is done by unions or by firms (who set efficiency wages). Both unions and employers have incentives to set wages above market clearing levels. Thus this labour market does not clear at a competitive-labour-market wage and employment level, leaving some workers involuntarily unemployed. They fall back on the secondary sector for employment (whilst perhaps continuing to search for primary sector jobs).
(b) *Secondary sector/segment***:** On both the labour demand and labour supply sides, the secondary sector economy faces obstacles and lacks security. Thus, the derived demand for labour is insecure and fluctuating – it is the less-secure jobs sector, also with much underemployment and survivalist-level activities. Those in self-employment in the secondary sector face similar insecure conditions. There also are distance, network, information and other barriers to entry, meaning that there are workers who not only fail to be employed in the primary sector, but also fail to enter the secondary sector. These workers remain involuntarily unemployed – ending up in the tertiary segment.
(c) *Tertiary segment:* This segment comprises those who remain unemployed. It is the default or last-resort segment, the *'no-jobs' segment*. There is no labour market or market-oriented self-employment to speak of. Livelihoods rely on savings, social or intra-family transfers, and so forth – whilst continuing to search for jobs, or alternatively becoming non-searching/discouraged (whilst remaining part of broader labour force).

This model is developed by augmenting the Layard *et al* model by including barriers to entry with regard to the secondary sector – and explicitly recognising the tertiary segment. Such labour market barriers will limit the effective supply of labour to the secondary sector. A lim-

[6] Otherwise-conceptualised segmentations (or sub-segmentations) could be entertained and probably be dealt with in similar ways.

ited labour supply causes the unemployment rate to be higher than what it would be in the absence of the barriers. Thus, unemployment in the augmented model will be higher than in the LNJ model, which in turn is higher than in the NCBM. Because (in the augmented model) these additional unemployed would have supplied their labour bar these barriers, they can be characterised as involuntarily unemployed. Thus, a key feature of the barrier-augmented model is that it demonstrates the possibility of the existence of involuntary unemployment in long-run equilibrium. These involuntarily unemployed constitute the tertiary segment.

5.2 The mechanics of the three-segment model in comparison to the NCBM and the LNJ model

This subsection develops the three-segment model sequentially with reference to two other 'benchmark' models, comparing the unemployment and wage levels that they imply. It does so in an adaptation, but also a deeper exploration, of the Layard *et al* (LNJ) two-sector model presented above. The primary sector mainly[7] represents firms (and unions) that behave largely as described by the New Keynesian models discussed above.The secondary sector does not display such behaviour, simply because its firms do not possess market power or face union power. Following Layard *et al.*, the secondary sector is defined also to include workers who are, or become, self-employed in less-secure form.

5.2.1 The impact of New Keynesian conditions in the primary sector on aggregate unemployment: a deeper look at the LNJ two-sector model

The first step is to reconstruct the LNJ model to analyse its characteristics more rigorously. As a point of reference, consider a perfectly competitive (NCBM) two-segment model of an economy. The wage level will be the same in both sectors, brought about by perfect labour mobility between sectors. In figure 6 this reference condition is shown by the intersection of the supply and demand curves for labour in respectively the primary and secondary sectors L^S_{PC} and L^D_{PC} for the primary sector and L^S_{SC} and L^D_{SC} for the secondary sector. The secondary sector is measured on the horizontal axis from zero on the right-hand side.

The equilibrium wage rate in this completely competitive (subscript C) model is $W_{PC} = W_{SC}$. At equilibrium (indicated by •), employment in the primary sector will be E_{PC} and in the secondary sector E_{SC}.

Unemployment will equal the distance represented by the double-headed arrow entitled *a* = the distance between E_{PC} and E_{SC}. All unemployment is voluntary. (The length of the horizontal axis indicates the entire labour force, i.e. the economically active. The horizontal

[7] Though, in reality large sections of the formal sector do operate in competitive or nearly competitive markets (e.g. professionals such as lawyers and doctors, or retail outlets of various nature), most commodities and services in the modern economy are traded in monopolistically competitive markets. Hence, the simplifying assumption in the models discussed below that the primary sector comprises monopolistically competitive firms.

axis has been extended to also show those who are not economically active, numbering N_N. They are outside the labour market.)

Next we introduce New Keynesian (subscript K) features to the primary sector, i.e. imperfectly competitive conditions. Accordingly, firms have price-setting power, while either firms or unions have wage-setting power. Therefore, the analysis uses the price-setting (PS) and wage-setting (WS) relationships for the primary sector, as before. In the primary sector labour market equilibrium (indicated by ■) will occur at the intersection of WS_{PK} and PS_{PK}, at wage W_{PK} and employment level E_{PK}. Compared to the NCBM equilibrium, a number of workers equal to the distance *b* will not be employed in the primary sector as a result of the presence of imperfectly competitive conditions and New Keynesian behaviour such as efficiency wages.

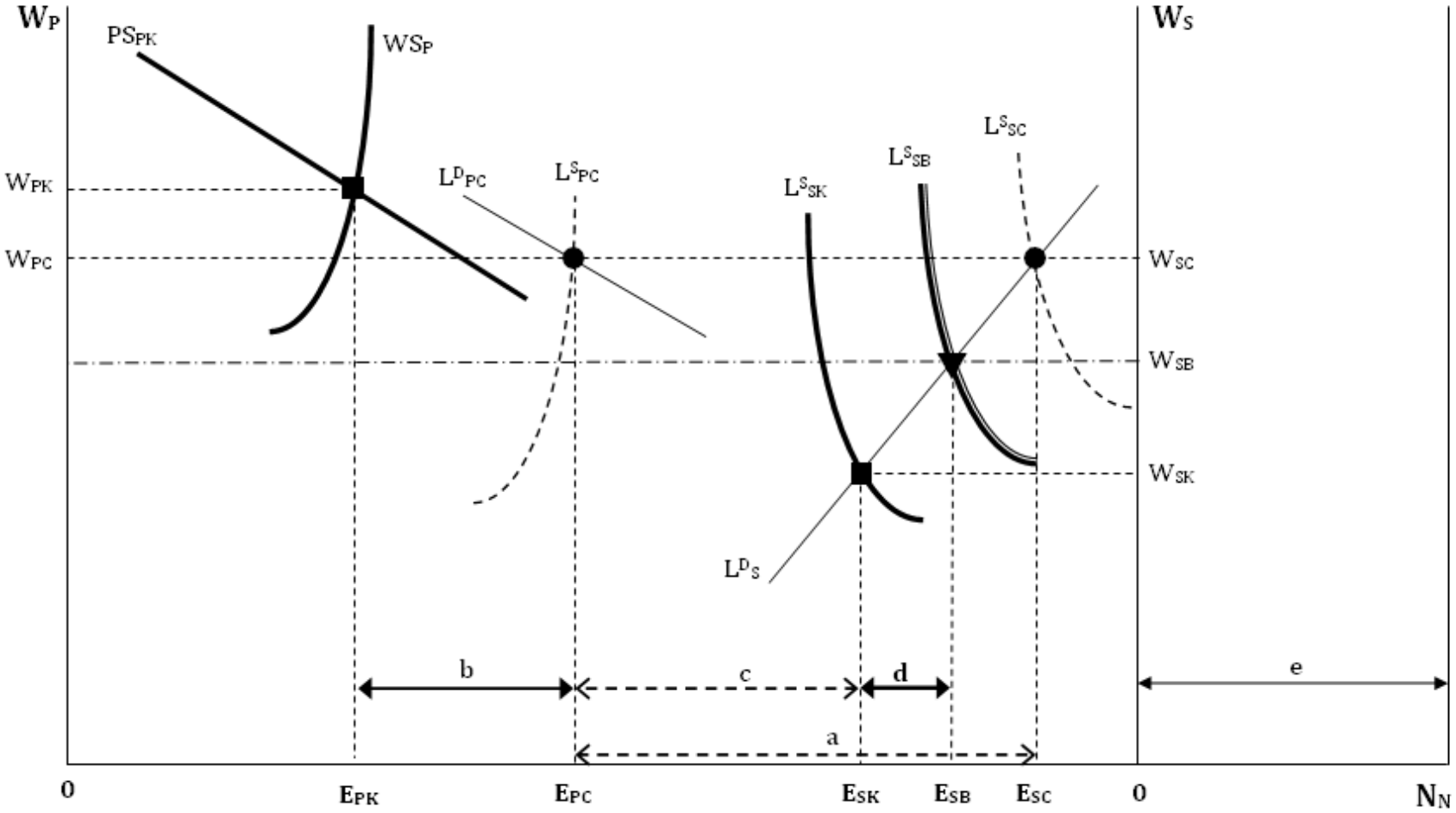
However, these workers might not be unemployed. They have two other options. They can be economically inactive, meaning that N_N, will be larger than in the NCBM reference case (the double-headed arrow *e* will lengthen). Or, they can work in the secondary sector, in which case labour supply in the secondary sector will be higher than in perfectly competitive reference case. Hence, labour supply will lie to the left of L^S_{SC} at L^S_{SK}. (Ignore L^S_{SB} for the moment.) If all those not employed in the primary sector (due to the presence of the New Keynesian features) work in the secondary sector, the horizontal displacement from L^S_{SC} to L^S_{SK} will equal distance *b*. However, because of the slopes of L^S_{SK} and L^D_S, employment in the secondary sector will be higher by less than *b*. Equilibrium in the secondary sector (the other ■) occurs at employment level E_{SK}. (This is equivalent to the result in figure 5.)

Unemployment in this case (which is the LNJ model) will be higher than in the NCBM case: *b* plus *c* in this model will exceed *a* in the NCBM. The reduction of labour absorption in the primary sector will exceed the gain in employment in the secondary sector.

Note that, although the primary sector is now characterised by imperfectly competitive conditions, the secondary sector is assumed to be still perfectly competitive (in keeping with the LNJ model). There is no price-setting and wage-setting behaviour in this sector.

Since distance *b* in figure 6 represents those who are not employed in the primary sector because of its imperfectly competitive conditions, and because these workers would be willing to supply their labour in the absence of these features, they might be termed “involuntarily unemployed in the primary sector”. This is the standard New Keynesian single-sector result.

Figure 6 – Unemployment in the three-segment 'barrier' model



If these workers do not become economically inactive, they supply their labour services in the secondary sector. Given that the secondary sector is assumed to be perfectly competitive, the residual, unabsorbed group who are unemployed (i.e. *b* plus *c*) are 'voluntarily unemployed in the secondary sector'. This also leaves them *voluntarily* unemployed in the final instance. In the aggregate there are no involuntarily unemployed left. *This is the basic LNJ two-sector* result – a perhaps surprisingly New Classical outcome. It also means that the tertiary segment in this case is empty – the model cannot explain the existence of persistent involuntary unemployment, which generates the tertiary segment.

In much of the developing economy, duel-economy literature the informal sector is seen in this way. It means that the informal sector is seen as soaking up everybody in the labour force who is not absorbed in the formal sector, in effect leaving only the voluntary unemployed – a weakness of that literature and of the two-sector LNJ model (also see, for instance, Bulow and Summers (1986) where the secondary sector merely absorbs those not employed in the primary sector).

It is to be noted that, even though aggregate unemployment in the LNJ two-sector model is voluntary, as a result of worker and employer preferences the unemployment rate in the model exceeds the unemployment rate that would exist in the absence of the imperfectly competitive conditions (New Keynesian features) of the model.

Also note that some individuals might also decide to leave the labour market altogether and join the economically inactive, thus increasing $N_{N,}$ and lengthening *e* and shifting the secondary sector vertical axis left, reducing *c*.

5.2.2 Augmenting the model and diagram: Introducing the impact of labour market barriers in the secondary sector

The discussion above highlighted several reasons why the secondary sector might be an imperfectly competitive sector. Several barriers to entry might exist for those wishing to work in the secondary sector. Since the secondary sector is defined to include some self-employed persons, these barriers also imply barriers to self-employment. Entry barriers mean that not all of those who cannot find employment in the primary sector will be able to supply their labour in the secondary sector. This also applies to individuals who have not been employed in either of the sectors, e.g. those in the tertiary segment.

Thus, barriers – for example, a lack of skills (i.e. human capital) or the basic physical capital needed to operate in the secondary sector – means that, the *effective* supply of labour in the secondary sector is less than the no-barriers labour supply or '*desired'* supply. In figure 6 this is represented by a labour supply curve, L^S_{SB}, that lies to the right of L^S_{SK}. Secondary sector

employment in the presence of barriers, E_{SB} (equilibrium indicated by ▼), will be lower than employment in the absence of the barriers, E_{SK}.

Note that L^S_{SB} represents a constrained supply curve, which means that the employment level at which it intersects the labour demand curve for the secondary sector, is off the desired (or what would have been the unconstrained) supply curve L^S_{SK}. The barriers themselves, as modelled here, operate in the secondary sector, but can be argued to originate in the financial markets. A lack of human or physical capital could, in principle, be overcome by borrowing the financial resources needed to create such capital (there might, of course, be other barriers too, but this article focuses on capital barriers). However, because of asymmetric information interacting with low returns associated with small-scale capital, financial institutions may face an adverse risk-return trade-off, rendering them very reluctant to extend loans to finance the creation of such human and physical capital.

The secondary sector wage in the presence of barriers will be at W_{SB}, higher than W_{SK} in the LNJ model, thereby, ironically, implying a smaller wage differential compared to the LNJ model. The secondary sector equilibrium wage will, however, always be below the primary sector wage.[8]

What about unemployment? First, the unemployment rate in the augmented model is higher than in the no-barriers/LNJ model. In the presence of barriers in the secondary sector unemployment will equal *b* plus *c* plus *d*, compared to only *b* plus *c* in the LNJ model. Secondly, in long-run labour market equilibrium a proportion of the unemployed are involuntarily unemployed. While *b* equals the number of workers who are 'involuntarily unemployed in the primary sector', and *b* plus *c* represents the number of 'voluntarily unemployed in the final instance', *d* represents the number of those who are, in the aggregate, 'involuntarily unemployed in the final instance'. In the absence of the barriers these workers would have been able to supply their labour to the secondary sector – they now constitute the tertiary segment, defined as the unemployed. Thus this barrier model (with barriers in the secondary

[8] W_{SB} in Figure 6 has been drawn to be lower than the wage in the primary sector. Should it be higher, one would expect workers employed in the primary sector to shift their supply to the secondary sector until wages are equalised. However, when wages are higher in the primary sector, because of the New Keynesian features present in the primary sector, workers cannot shift their supply from the secondary to the primary sector. Hence, wages in the secondary sector will always be lower or equal to wages in the primary sector.

sector[9]) explains the existence of a tertiary segment containing persons who are, in the final instance, involuntarily unemployed persons in long-run labour market equilibrium.[10]

Although it might be unusual to present an analysis in which involuntary unemployment is present even though the labour market is in equilibrium (given that its two subsectors are in equilibrium), it should be noted that the equilibrium in the secondary sector labour market is a *constrained* equilibrium. As already argued (and discussed further below), the constraint originates in a problem in the credit market where individuals are unable to borrow the funds needed to obtain the physical and human capital required to operate in the secondary sector.[11]

5.3 Dynamics in the three-segment model: shocks and hysteresis

The previous subsection compared the NCBM, LNJ and three-segment, "augmented LNJ" models in long-run equilibrium. An important result in the three-segment model with labour market barriers was the presence of higher total unemployment and of involuntary unemployment in long-run equilibrium.

This section discusses the effect of a demand shock and the sequence of events that follow if hysteresis occurs. We first consider the model without labour market barriers.

5.3.1 Hysteresis but no labour market barriers

While hysteresis can be introduced to both sectors, we will first introduce it to the primary sector and then to the secondary sector. As before we use the LNJ model as the point of departure. In figure 7 the model starts with a primary sector equilibrium at the intersection of WS_{PK} and PS_{PK} and a secondary sector equilibrium at the intersection of L^{S}_{SK} and L^{D}_{S}.

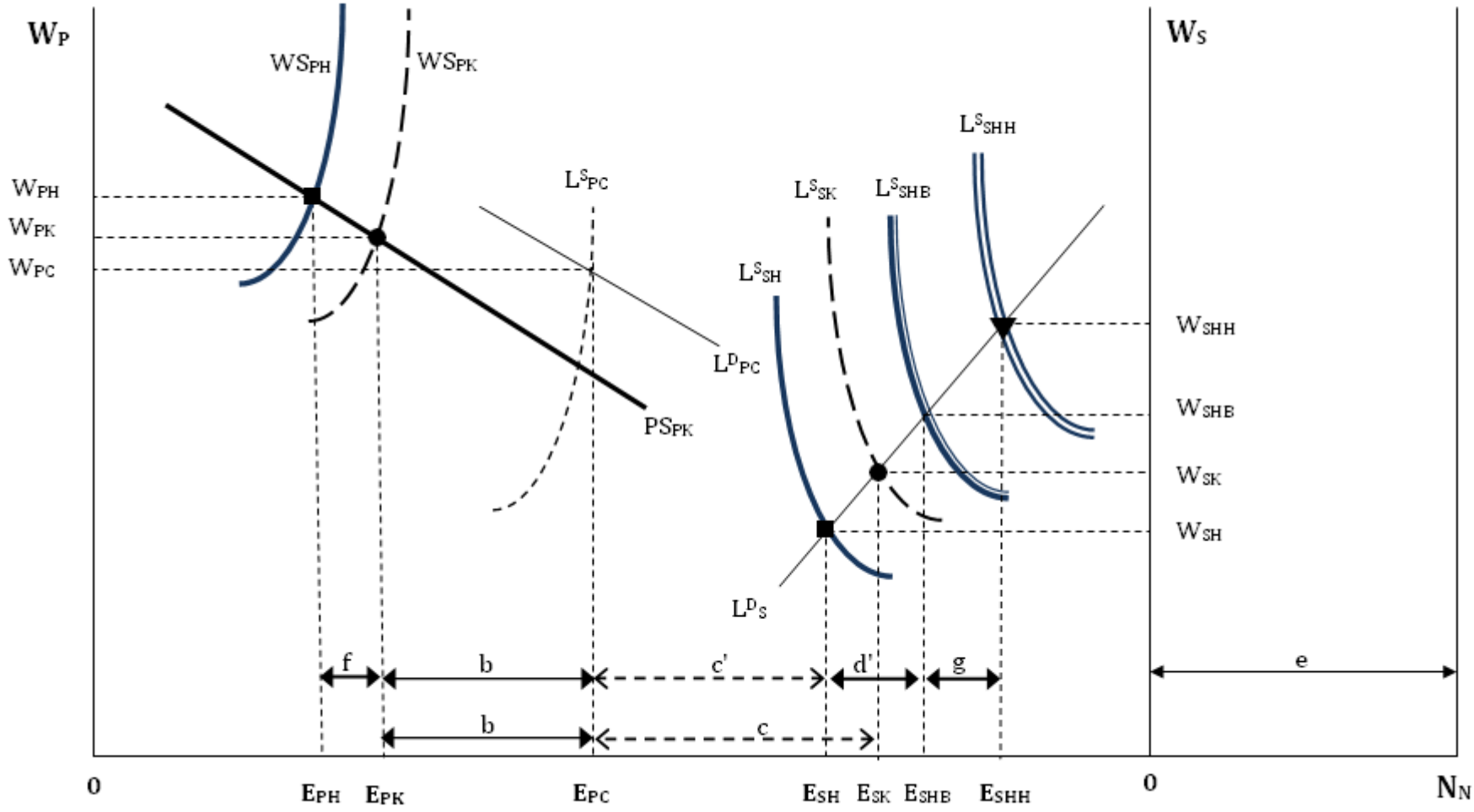
Should a demand shock affect the primary sector – contracting output, labour demand and employment in the short to medium run – and should those who become unemployed as a result of the demand shock become unemployable in the primary sector due to e.g. a loss of

[9] One can also introduce labour market entry barriers in the primary sector. It would analogously reduce effective labour supply in the primary sector and produce a higher wage and a lower level of employment in equilibrium. More workers would be excluded from this sector; resultant impacts on the secondary and tertiary segments follow logically. Importantly, this also means that there will be more persons involuntarily unemployed in the primary sector. (At least some of them will also face a barrier to the secondary sector, thus pushing them into the tertiary segment.)

[10] Some individuals might also decide to leave the labour market and become economically inactive, thus increasing $N_{N,}$ lengthening *e,* shifting the secondary sector vertical axis left and reducing *c,* the size of the tertiary segment.

[11] In the literature the work by Clower and Leijonhufvud serves as precedent for a model that has a constrained equilibrium at which involuntary unemployment exists. In these models transactions that fail to occur in one market constrain effective demand in another market (Leijonhufvud 1967:402-3) – as Clower (1965) famously suggested, if he fails to sell his consulting services in the labour market, it will constrain his demand for champagne in the goods market. Similarly, in the model discussed below problems in the financial market in obtaining finance for physical and human capital needed to operate in the secondary sector, limit effective labour supply in the secondary market.

Figure 7 – The three-segment 'barrier' model with hysteresis



skills, we have hysteresis: WS moves from WS_{PK} to $WS_{PH.}$ The long-run employment equilibrium (in the primary sector) shifts to a lower level and remains stuck at E_{PH}.

If these unemployed move their supply of labour to the secondary sector, L^S moves from L^S_{SK} to L^S_{SH}. Whereas unemployment equals *b* plus *c* in the LNJ New Keynesian model, it will equal *b* plus *c*' plus *f* in the LNJ New Keynesian model with hysteresis effects. If all those who become unemployed as a result of hysteresis shift their labour supply to the secondary sector (they could also leave the labour market), then the horizontal (leftward) shift from L^S_{SK} to L^S_{SH} will equal *f*.

The increase in employment in the secondary sector will be less than *f* (assuming that the supply and demand of labour in the secondary sector respectively have a positive and negative slope). Thus, the reduction in employment in the primary sector due to hysteresis effects is not offset by the increase in employment in the secondary sector.

Therefore, the total number of unemployed in the LNJ-type segmented model with primary-sector imperfectly competitive conditions *and* hysteresis effects (i.e. *c*' plus *b* plus *f* in figure 7) will exceed the number of total unemployed in the LNJ New Keynesian model (i.e. *b* plus *c* in figures 6 and 7), which is already larger than the number of unemployed in the New Classical reference case (i.e. *a* in figure 6). Similarly, involuntary unemployment in the primary sector will equal *b* plus *f*, which is higher than in the other cases. In the aggregate, involuntary unemployment 'in the final instance' still is zero (because at this stage the secondary sector still is assumed to be market-clearing).

5.3.2 Hysteresis plus labour market barriers

If the secondary sector is characterised by barriers to entry, then the effective supply of labour in the secondary sector[12] will be constrained (say at L^S_{SHB}) and therefore less than the desired supply of labour (L^S_H). Employment in the secondary sector will be at E_{SHB}, meaning that a number of workers equal to *d*' in figure 7 will be added to the unemployed already identified (thus totalling *d*' plus *c*' plus *b* plus *f* in figure 7). They (*d*') will be involuntarily unemployed in the secondary sector, and *d*' then also is the total involuntarily unemployed in the final instance.[13] This completes the analysis of hysteresis in the primary sector. Note that it affects both sectors.

[12] The barriers will constrain both the labour supply in the secondary sector as such and the absorption of unemployed individuals who flow there from the primary sector due to hysteresis. Effective labour supply L^S_{SHB} is drawn to show the total effect of the barriers, i.e. after some overflow from the primary sector.

[13] Given that in the LNJ-with-hysteresis model those workers not accommodated in the secondary sector now comprise both those who did not get employed because of the New Keynesian features of the primary sector (i.e. concentrated goods markets, efficiency wages and unions operating to ensure higher wages to their members) as well as those who lost their jobs in the primary sector because the shock, one might expect d' in the LNJ-with-hysteresis model (figure 7) to exceed d in the LNJ-without-hysteresis model (figure 6).

If (in addition to barriers to entry) hysteresis also occurs in the secondary sector, a demand shock will cause persistent/long-run unemployment in the secondary sector. Unlike the primary sector workers who moved to the secondary sector upon losing their jobs, job-losers in the secondary sector have no other employment sector to move to. Their options are to remain (involuntarily) unemployed (i.e. to enter the tertiary segment) or move out of the labour force. If their continued unemployment causes these workers to lose relevant skills, the labour supply in the secondary sector will decrease and L^S will move from L^S_{SHB} to L^S_{SHH}; equilibrium employment will be at E_{SHH}. Total as well as involuntary unemployment will increase with a further number of workers equals to *g*, bringing the total involuntarily unemployed in the final instance to *d' plus g* in figure 7. The tertiary segment will be significantly larger than in any previous case. (The number of purely voluntarily unemployed is *c*'. Recall that *b* plus *f* represents those 'involuntarily unemployed in the primary sector', but 'voluntarily employed with regard to the secondary sector'. Thus those in *b* and *f* are not part of aggregate involuntary unemployment)

Therefore, if hysteresis is present in both sectors, a demand shock will detrimentally impact on equilibrium employment in both the sectors. The secondary sector is affected twice, first through the spill-over of individuals from the primary sector and secondly through the direct impact of the shock on the secondary sector.

With hysteresis in both sectors, the post-shock long-run equilibrium will exhibit higher total unemployment and higher involuntary unemployment 'in the final instance' than without hysteresis.

5.4 Main results of the conceptual-diagrammatical three-segment barrier model

- In a basic two-sector segmented model with New Keynesian features in the primary sector, in the aggregate there is no persistent involuntary unemployment. (This is the basic LNJ two-sector result.) Such a model does not explain the existence of a tertiary segment, defined as comprising the involuntarily unemployed.
- If there are labour market barriers in the secondary (and/or primary) sector, the model explains the presence of higher total unemployment and specifically of involuntary unemployment in long-run equilibrium. Thus, the barrier model explains the existence of a tertiary segment that includes the involuntarily unemployed.
- If hysteresis is present in either sector (or in both), it results in higher unemployment and, specifically, involuntary unemployment 'in the final instance' in long-run labour market equilibrium.
- The presence of both labour market barriers and hysteresis results in even higher unemployment and, specifically, higher involuntary unemployment 'in the final instance' in

long-run labour market equilibrium – and a larger tertiary segment than with only one of these phenomena.

The key feature of the three-segment barrier model is that the secondary sector cannot just adjust its wages/earnings to soak up all of the unemployed looking to be employed or self-employed. Thus, given the nature of bargaining and competition (or the lack thereof) in the primary sector (i.e. given that both unions and employers have incentives to set wages above market clearing levels), the inability of the secondary sector to absorb all those who cannot find employment in the primary sector means that the economy will be left with involuntary unemployment.

6. A mathematical three-segment barrier model

Having outlined a conceptual, diagrammatical three-segment model in section 5 and derived its results above, we now develop a mathematical three-segment model for an economy such as that of South Africa. (It draws on the prior diagrammatical exposition, but is not precisely equivalent in all respects as it makes a few simplifying assumptions along the way.)

This section of the paper draws on the dual labour market model of Bulow and Summers (1986), which itself is an augmentation of an efficiency wage model – a prominent approach in the New Keynesian class of models. Bulow and Summers uses an efficiency wage model to explain why some workers are not accommodated in the primary sector; however, those workers all find employment in a secondary sector. Their dual labour market model explains the *allocation* of workers between the primary and secondary sectors – but not the existence of involuntary unemployment.

To augment this approach, and following Summers (1988) as well as Knell (2014), Pereau and Sanz (2006), Bulkley and Myles (1996) and the suggestion by Bulow and Summers (1986), the paper introduces union bargaining into the model to allow for the presence of strong labour unions in the South African economy. Similarly, the presence of high economic concentration and imperfectly competitive product market conditions is an integral part of our augmented model.

Furthermore, the model also takes account of a typical South African feature of the labour market resulting from the spatial dimensions of the policy of apartheid. During the apartheid era, black townships were located some distance from business centres, while many so-called black homeland areas were established in remote areas of the country, with black workers required to obtain passbooks if they wanted to work outside these homelands (through a system called ‘influx control’). Even though it has been two decades since apartheid has been abolished, spatial patterns persist, raising the travel cost of looking and holding a job. The model incorporates this element.

In contrast to the dual labour market model of Bulow and Summers (1986) that merely explains the allocation of workers between the primary and secondary sectors, but not the existence of unemployment, the model in this paper incorporates Kingdon and Knight's (2004) and Grimm, Krüger and Lay's (2011) suggestion that workers end up being unemployed and not in informal sector employment because of barriers to entry into the informal sector. More specifically, the mathematical model will incorporate financial barriers arising from a lack of financing for high transport costs (and the long distances to and from places of work) that prevent some workers from searching for jobs in the primary sector. Secondly the model will incorporate barriers to entering the secondary sector that result from a lack of funding to set up the (physical and human) capital needed for secondary sector activity. As will be shown, the presence of these barriers implies the existence of a third segment that comprises the unemployed (who survive in various ways).

6.1 Key characteristics of the South African labour market: implications for the macro model

As noted above, the following 'stylised facts' characterise the South African economy and labour market and should feature in the macro model:

1) A distinction, or segmentation, between those workers and self-employed persons in secure and stable jobs, typically in established, registered companies, and those in 'not so good' and rather unstable jobs, often in micro-enterprises that are not formally registered, including informal self-employment (as owners of informal one- or multi-person enterprises) – i.e. a primary and a secondary sector (compare section 5.1).
2) Economic concentration and the presence of imperfectly competitive behaviour such as monopolistic price setting and wage setting in the primary sector, as well as the possibility of efficiency wages.
3) A system of central wage bargaining in most formal industries with strong unions (with firms operating in highly concentrated industries on the other side).
4) A secondary sector which is not concentrated, but has barriers to entry for prospective employees and self-employed persons.
5) A high rate of long-term, open unemployment.

Stylised facts 1-3 will be used as assumptions in the model, while the model sets out to explain stylised facts 4 and 5. We deal with the model in two phases: first, a two-sector model without involuntary unemployment (section 6.2 below) as a preparatory step towards an expanded two-sector, *three-segment* model with involuntary unemployment (section 6.3).

Using point 1 above, we draw on the dual labour market literature to present, as a starting point, a two-sector model with a primary and a secondary sector. The primary sector comprises firms that operate and set product prices in a monopolistic-competition context.

Efficiency wages are paid in an attempt to ensure productive work effort of workers (and discourage low-productive work, or shirking). The primary sector is also a sector with desired, or 'good', secure and stable jobs.

The payment of efficiency wages means that there are workers who will be involuntary unemployed in the primary sector (i.e. their reservation wage is equal to or lower than the market wage, yet they are unable to find employment in the primary sector). In the initial two-sector model (section 6.2), workers who are involuntary unemployed in the primary sector will find employment in the secondary sector. The secondary sector is characterised by jobs that are less attractive than those in the primary sector – it has 'not so good', less-secure and unstable jobs. Firms in the secondary sector usually are significantly smaller than in the primary sector; therefore, effort will be assured without the need to pay efficiency wages.

In addition to the payment of efficiency wages by primary-sector firms, the model includes labour union bargaining behaviour, thereby also accounting for points 2 and 3 above. This combination is in line with Bulow and Summers (1986:377, fn 1), who argue that non-shirking models can encompass labour union behaviour too, as the presence of labour unions may render the reduction of wages by firms expensive. Efficiency wages and labour union bargaining becomes only operative in monopolistically competitive markets – thus, the model also incorporates an assumption that the primary sector comprises monopolistically competitive firms. Summers (1988), Knell (2014), Pereau and Sanz (2006), and Bulkley and Myles (1996) all show how an efficiency wage model with its non-shirking component can be combined with a labour union model.

In the second phase (section 6.3 below) we address points 4 and 5 above. We expand the two-sector model (which has no involuntary aggregate unemployment) to a two-sector three-segment model *with involuntary unemployment*. This model allows for workers to be involuntary unemployed in both the primary and secondary sectors and therefore to be involuntary unemployed on the aggregate level. The model achieves this by also introducing the entry barriers into the informal sector discussed above.

6.2 Step 1: The two-sector model with no involuntary aggregate unemployment

We derive a formal-sector job-offer relationship and an effort supply function.[14] Different from the analysis in Bulow and Summers (1986), this analysis is done in terms of the number

[14] Concerning the microfoundations of the model, the model assumes a simple utility function, resembling the specification by Bulow and Summers (1986), with infinitely lived agents, where utility, *Ut*, is a function, *f*, of consumption and shirking (or 'non-effort'):

$$Ut = f(x_p, x_s + \alpha l)/r \qquad (1)$$

where x represents consumption of goods produced in the primary and secondary sectors (subscripts p and s denote the primary and secondary sectors). In addition, l is zero when the worker exerts effort and one if the

of *positions* filled by firms rather than the number of workers demanded, which allows the introduction of factors that will influence the number of positions being filled by firms in the two sectors. Nevertheless, the model is presented in terms of both the number of positions and the positions filled (persons employed).

In addition to these two relationships, the analysis below also presents wage-setting and price-setting relationships. These four relationships are then used to derive equilibrium conditions for the primary and secondary sectors.

6.2.1 The effort supply function

At any given moment firms in the primary sector fill a number of positions (jobs). The total number of jobs available in the primary sector is F_p. Those workers who do not obtain employment in the primary sector are accommodated in the secondary sector (which is assumed to be without entry barriers). In the secondary sector there is equilibrium: the total number of jobs filled is F_s. Thus, although there might be involuntary unemployment in the primary sector, there will not be involuntary unemployment at the aggregate level. The total number of filled positions in the economy (which in this case amounts to the entire labour force) is $F = F_p + F_s$; the allocation between the two sectors can be described in terms of the proportion of total positions filled by firms in the primary sector being $p = F_p/F$, while the proportion filled by firms in the secondary sector is $(1 - p) = F_s/F$.

A worker who quits or is laid off in the primary sector, is assumed to move to the secondary sector. The quit rates in the primary and secondary sectors are q_p and q_s; d_2 represents the probability of the worker being laid-off when caught shirking (or for e.g. low productivity[15]), while d_1 represents the probability of being laid-off for shirking while not actually shirking (a false positive). Furthermore, w_p and w_s represent the wage rates in the primary and secondary sectors. Therefore, $(1 - q_p - d_1)w_p$ represents the expected wage of those workers employed in the primary sector (i.e. who have not been laid-off and have not quit the primary sector), while $(q_p + d_1)w_s$ represents the expected wage of primary sector workers who are laid-off in or quit from the primary sector and move to the secondary sector. (Shirkers are assumed to produce nothing, hence their $PV = 0$ and they are not included.) Likewise, $(1 - q_s)w_s$ represents the expected wage of those workers in the secondary sector who remain in the

---

worker does not exert effort. Non-effort is thus considered to be a consumption good, and it is substitutable for secondary sector goods. Furthermore, α is the instantaneous gain in utility from shirking/non-effort, while r represents the discount rate. Following Bulow and Summers (1986) we assume risk neutrality (so that $f(\lambda x_p, \lambda x_s) = \lambda f(x_p, x_s)$) and preferences are homothetic and normalised (so that $f(0,0)=0$).

[15] For simplicity, quitting and being laid off are modelled to depend on shirking (insufficient work effort/productivity); other factors that determine quitting or being laid off can be modelled analogously. The simplification is not central to the main result of involuntary unemployment present in the full model, but merely facilitate it – involuntary unemployment will depend on the presence of barriers to entry into the secondary sector. Nevertheless, because it is commonly used in international literature, the shirking model it is used here.

secondary sector, while $q_s w_p$ represents the expected wage of those workers who quit the secondary sector for the primary sector. Thus, the sum of the present value of expected primary and secondary sector income in the economy is:[16]

$$PV=[(1 - q_p - d_1)w_p/r + (q_p + d_1)w_s/r]p + [(1 - q_s)w_s/r + q_s w_p/r](1 - p) \quad (2)$$

In equilibrium, labour flows into and out of the primary sector need to be equal. Thus $p(q_p + d_1) = q_s(1 - p)$, so that $q_p + d_1 = q_s(1 - p)/p$. This equality also means that search for work in the primary sector occurs not from a position of unemployment, but from the secondary sector (in the two-sector model there is no aggregate unemployment).

Following Bulow and Summers (1986), we define an *effort supply function*. The effort supply function is stated in terms of α, defined as the instantaneous gain in utility from not exerting effort, as follows:

$$\alpha \leq (d_2 - d_1)(PV_p - PV_s) \quad (3)$$

where $(d_2 - d_1)(PV_p - PV_s)$ represents the gain from non-shirking/effort; PV_p is the present value of primary sector work and PV_s the present value of secondary sector work (recall that non-effort is only possible in the primary sector, the sector that pays a wage premium over the secondary sector wage). This conditional expression shows the premium that firms pay (the right-hand side of equation 3) to overcome the gain that workers derive from not exerting effort (the left-hand side of equation 3), thereby ensuring that they exert effort.

As mentioned above, the model in this paper combines an efficiency wage model (with its non-shirking component) with a labour union model. As a result α includes also the premium that companies have to pay to ensure the effort of unionised labour (i.e. to ensure that unionised workers limit their strike action or do not strike at all). This will render $\alpha = \alpha_1 \alpha_2$, where α_1 is the instantaneous gain in utility from not exerting effort (i.e. from shirking), and α_2 (which is > 1) constituting the premium that unionised workers can extract.[17, 18]

[16] For reasons of simplicity equation 6 assumes infinitely lived workers and as such uses the simple formula for the calculation of the value of a consol to calculate the present value.

[17] The mark-up/premium rate is ($\alpha 2 - 1$).

[18] The South African labour market is also characterised by a clear skills-related stratification of the unemployed, with an oversupply of unskilled workers and a shortage of skilled workers: the unemployment rate among individuals holding post-school degree qualifications is approximately 5%, and among those who have not completed school just below 50% (CDE 2013; Van der Berg and Van Broekhuizen 2012). This paper does not include these highly skilled workers into the model simply because when they quit or are laid-off they typically do not move to the informal sector, but find employment relatively easily elsewhere in the formal sector. Highly skilled workers will probably also be able, given the tightness of their submarket for labour, to negotiate a premium on their income. For these workers a matching model for hires of high skilled workers, η, could be used, where a scarcity of skilled workers would explain a low unemployment rate of skilled workers,

The South African labour market is also characterised by significant spatial distortions resulting from Apartheid, where places of residence of black people very often were far removed from places of work (in the primary sector). These distances significantly raise travel costs, which need to be added to the premium that workers require before working in the primary sector. Therefore:

$$\alpha = \alpha_1\alpha_2 + \alpha_3 D \qquad (4)$$

where D represents the distance between place of residence and place of work in the primary sector, and α_3 represents the cost per unit of distance.

1. Unions having more power implies a higher value of α_2 and therefore a higher value of α; consequently, the difference between the present values of primary and secondary sector wages will be higher.

2. Similarly, the larger α_3 and D, the larger will be the value of α. The inclusion of the term $\alpha_3 D$ means that both distance and the cost per unit of distance impacts the reservation wage of workers – and negatively affects job search. If people live far from places of work in the primary sector and have to travel to places of work, they may not be able to afford job search.

   Note that this particular search/entry barrier can be seen as principally due to a financial market failure. Jobseekers find it hard to borrow money to finance for their traveling and search costs (intending to repay the loan upon finding a job). Lenders might be unwilling to extend such loans due to both a low probability of finding a job and a low expected wage.

Rearranging equation 3:

$$\alpha/(d_2 - d_1) \leq (PV_p - PV_s) \qquad (5)$$

Using equation 2, the present values of being employed in the primary and secondary sectors are:

$$PV_p = [(1 - q_p - d_1)pw_p + q_s(1 - p)w_p]/r$$

$$PV_s = [(q_p + d_1)pw_s - (1 - q_s)(1 - p)w_s]/r \qquad (6)$$

$U_{SKILLED}$, and a high vacancy rate of skilled people, $V_{SKILLED}$, and where such scarcity can also explain a high wage premium. The hire rate model would be: $\eta_{Skilled} = \beta m(U_{SKILLED}, V_{SKILLED})$.

Therefore, using equation 6:

$\alpha r/(d_2 - d_1) \leq (1 - q_p - d_1)pw_p + q_s(1 - p)w_p - ((q_p + d_1)pw_s + (1 - q_s)(1 - p)w_s)$

which reorganises as:

$\alpha r/(d_2 - d_1) \leq ((1 - q_p - d_1)p + q_s(1 - p))w_p - ((q_p + d_1)p + (1 - q_s)(1 - p))w_s$

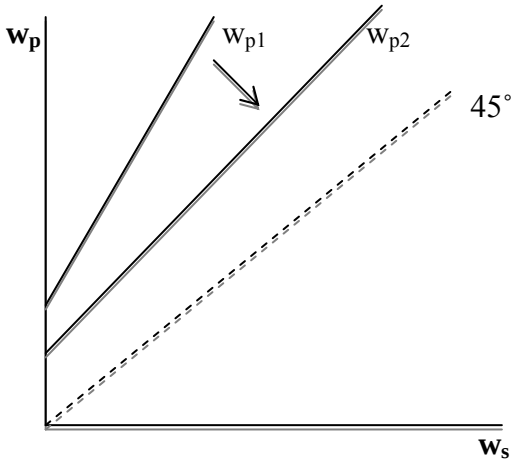
and after normalising on w_p yields:

$w_p \geq \alpha r/(d_2 - d_1)((1 - q_p - d_1)p + q_s(1 - p)) + (((q_p + d_1)p + (1 - q_s)(1 - p))/((1 - q_p - d_1)p + q_s(1 - p)))w_s$ (7)

Recalling that $q_p + d_1 = q_s(1 - p)/p$ and substituting into equation 7 yields:

$w_p \geq \alpha r/(d_2 - d_1)p + (1/p)w_s$ (8)

Figure 8 – The relationship between primary and secondary sector wages



Equation 8 represents the effort supply function (equations 3 and 5 above) in a different form that shows the relationship between the *wage* and the *proportion of positions filled* in the primary sector: as p increases, w_p decreases. It also expresses the primary-sector wage as the secondary-sector wage plus a mark-up. (It still is an effort supply function: the mark-up or premium is what needs to be paid to primary sector workers to ensure effort.) Thus, the relative proportion of positions allocated to primary sector jobs (p) has an impact on the size of the mark-up on the secondary-sector wage rate. This is shown graphically in figure 8.

Note that, as p increases the slope of the relationship becomes flatter, while the intercept decreases (i.e. as p increases, w_p shifts and rotates from w_{p1} to w_{p2}).

6.2.2 *The price-setting relationship*

To derive the price-setting relationship we use the standard textbook equation stating the relationship between wages, the marginal product of labour (and hence the level of employment E) and profit mark-up of a monopolistically competitive firm. In equation 9 this is applied to the primary sector wage:

$$w_p = ((\varepsilon - 1)/\varepsilon)(MPL) = ((\varepsilon - 1)/\varepsilon)(b(E_p)) \text{ with } b' < 0 \text{ and } w_p > 0' \qquad (9)$$

with MPL being the marginal product of labour and ε the elasticity of product demand in a monopolistically competitive market (thus $(\varepsilon - 1)/\varepsilon < 1$, where $\varepsilon > 1$ to ensure that firms make a profit). MPL is defined as a negative function, b, of primary sector employment, E_p. Thus, holding ε constant, the primary sector wage becomes a negative function, g, of primary sector employment:

$$w_p = \gamma(b(E_p)) = g(E_p) \text{ with } g' < 0, \gamma = (\varepsilon - 1)/\varepsilon \text{ and } w_p > 0 \qquad (10)$$

where the size of γ relates to the size of the mark-up of a monopolistically competitive form; the higher γ and therefore the closer it moves to 1 (i.e. the closer ε moves to infinity and therefore approaches the perfectly competitive model), the lower can the mark-up be and the less the firm can benefit from its monopolistically competitive position.

Equation 10 represents the standard primary-sector *price-setting relationship* linking employment and wages: given that $g' < 0$, w_p decreases as E_p increases (but the wage cannot turn negative).

6.2.3 *The job-offer relationship*

The number of positions (F_p) and hence also the proportion of jobs/positions offered by firms in the primary sector, p, is a positive function of the marginal product of labour, which itself is a negative function of the level of employment (see the discussion of equations 9 and 10 above). Suppose, for reasons of simplicity, that this relationship is linear with parameter h:[19]

$$p = (h/\gamma)w_p = (h/\gamma)g(E_p) \qquad \text{or} \qquad w_p = p\gamma/h \qquad (11)$$

Thus, at higher levels of E_p the real wage is lower (because the marginal product of labour is lower), and hence, so is the proportion of positions filled by firms in the primary sector, p. Of course, if, for a given level of employment, the marginal product of labour increases – for

[19] Note that h is divided by γ so as to ensure that in equation 11 p is purely a function of the marginal product of labour and not γ: $w_p = g(E_p) = \gamma(b(E_p))$, so dividing $g(E_p)$ by γ leaves $b(E_p) = MPL$.

instance, due to an upgrade in skill levels – the number of positions offered in the primary sector will increase. Thus, the positive sign of h means that if workers are more productive, more workers can be employed at a given wage.

Given the role of the marginal product of labour in equation 11 and its link to the proportion of positions offered, equation 11 is also a *job-offer relationship* – it links the proportion of jobs/positions being offered to wages. (Below it will interact with equation 8, the effort supply function, to establish the equilibrium wage and number of positions filled.)

- Note that in terms of equations 10 and 11 there is a positive relationship between p and w_p, but a negative relationship between E_p and p (given that $g' < 0$): as E_p increases, w_p decreases, causing p to also decrease.

6.2.4 The wage-setting relationship

Substituting equation 11 into equation 8 yields equation 12:

$$w_p \geq (\alpha r/(d_2 - d_1)((h/\gamma)g(E_p)) + (1/((h/\gamma)g(E_p))w_s \qquad \text{with } g < 0 \qquad (12)$$

Equation 12 is a primary sector *wage*-setting equation with its characteristic positive relationship between the level of employment and wages. As E_p increases (and given that $g' < 0$), w_p increases simply because as employment in the primary sector increases (and hence, as employers offer more jobs), workers can get work easier elsewhere in the primary sector (the probability of getting a job in the primary sector is larger if a larger proportion of total jobs are filled in the primary sector) – hence firms need to offer a higher wage to ensure that they stay, exert effort and do not strike.

Equation 12 interacts with equation 10, the price-setting relationship between wages and employment, to determine the equilibrium values of wages and employment in the primary sector.

Workers in the secondary sector are just paid their marginal product, which, for simplicity, is assumed to remain constant: with little capital and similar skills and each person more or less operating on their own, they are assumed to have the same marginal productivity.

6.2.5 Model summary

The model can be summarised as follows.

First, in p-w_p space there are two relationships (the [ ] indicates the sign of the p-w_p relationship):

A job offer relationship

$p = (h/\gamma)w_p$ or $w_p = p\gamma/h$ [+] (11)

and an effort supply function

$w_p \geq \alpha r/(d_2 - d_1)p + (1/p)w_s$ [–] (8)

Secondly, in E_p-w_p space there are two relationships (with $g' < 0$) (the [ ] indicates the sign of the E_p-w_p relationship):

A price-setting relationship

$w_p = g(E_p)$ [–] (10)

and a wage-setting relationship

$w_p \geq (\alpha r/(d_2 - d_1)((h/\gamma)g(E_p)) + (1/((h/\gamma)g(E_p))w_s$ [+] (12)

Equations 8 and 11 – *or* equations 10 and 12 – can be used to calculate the equilibrium values of w_p, and equations 8 and 11 to calculate the equilibrium values of p. The expressions for w_p and p are:

$w_p = (\alpha r\gamma/(d_2 - d_1)h + w_s\gamma/h)^{0.5}$ (13)

$p = (\alpha rh/(d_2 - d_1)\gamma + w_s h/\gamma)^{0.5}$ (14)

To calculate the equilibrium value of E_p note that in equilibrium $E_p = F_p$ and that $p = F_p/F$. So, using equation 14 and given the value of F, equation 15 then produces the equilibrium value of E_p:

$E_p = (\alpha rh/(d_2 - d_1)\gamma + w_s h/\gamma)^{0.5}F$ (15)

Together with the effort supply function 8, the job offer relationship 11 then determines the equilibrium number of positions in the primary sector. Since the proportion of filled positions in the secondary sector is $(1 - p)$, the secondary sector absorbs all those who are not employed in the primary sector and who are willing to work for wage w_s. (This assumption will be relaxed in the next section). Thus, in this model – as in the model of Bulow and Summers – there is no involuntary unemployment.

6.3 Step 2: The two-sector, three-segment model with involuntary aggregate unemployment

In this section the model is expanded to contain a third sector/segment that comprises the unemployed. The preference hierarchy follows the model above: workers in the secondary sector prefer the primary to the secondary sector; the unemployed would prefer secondary sector employment to unemployment and primary sector employment to secondary sector employment.

6.3.1 The effort supply function

As in the previous section, we first consider the effort supply function. The effort supply function introduces a role for entry barriers that imply that not all of those who are unable to find a job in the primary sector will be able to find one in the secondary sector.

The model makes a few simplifying assumptions. First, those quitting and being laid-off in the primary sector (at rate $q_p + d_1$), move to the secondary sector, while those quitting the secondary sector (at rate q_s) move to unemployment (i.e. nobody moves from the secondary to the primary sector). Those of the unemployed who quit their unemployed status (at rate q_u) move either into the primary or the secondary sector. The unemployed, of course, receive no wage.

As before, the proportion of filled positions (jobs) supplied in the primary sector is p_p, while that of the secondary sector is p_s. A critical difference is that, unlike the two-sector model with no unemployment (where everyone who is willing to work in the secondary sector for a wage equal to their marginal product, w_s, finds employment), in this model the number of filled positions in the secondary sector, p_s, is equal to or less than $(1 - p_p)$; p_s being smaller than $(1 - p_p)$ would result from barriers to entry into the secondary sector. The barriers and obstacles may include physical, financial, human and social capital requirements.

Grimm, Krüger and Lay (2011) present a small model in which the barrier to entry results from the borrowing constraint of the potential secondary sector entrant interacting with the minimum scale of capital, K^*, needed to generate a higher return. Note that the capital, K, typically includes physical capital, but the concept can also be expanded to include human capital (i.e. the basic education and training needed to be employed by or operate a small enterprise). Thus, below the minimum scale the return to capital is very low. The question a potential entrant into the secondary sector faces is whether or not the minimum scale of capital is lower than her borrowing constraint. The borrowing constraint originates from asymmetric information: lenders do not know whether borrowers will in fact acquire the capital with their borrowed funds and thus be in a position to generate a return in excess of what the borrower needs to pay the lender for the borrowed funds. Thus, if the borrowing

constraint is lower than the minimum scale, then the return to capital is small, and the entrant will have to use her total return to cover the cost of capital, r_K; there will be no profit left after paying the cost of capital. Hence, investment will not take place and the entrant will not enter the secondary sector. If, however, the minimum scale is lower than the borrowing constraint, investment will take place and returns to capital will exceed capital cost (this high return will of course fall to zero as the scale of capital is expanded and the marginal product falls with the expansion in scale). In their model (Grimm 2011:S30) the secondary market entrant would maximise her profit, π, subject to a borrowing constraint, with output produced by a simple production function where $y = f(K)$, yielding output y produced with capital K when $K > K^*$, and capital producing just enough output to cover its cost when $K \leq K^*$:

$$\text{Max. } \pi = y - rK \qquad (16)$$

$$\text{s.t.} \quad y = f(K) \text{ if } K > K^*$$
$$y = r_K K \text{ if } K \leq K^*$$
$$\text{and } K \leq B^* \qquad (17)$$

The capital stock is chosen so that $f'(K) = r$ if $B^* > K^*$. If $B^* \leq K^*$, i.e. the borrowing constraint is binding, then the entrant is indifferent between different levels of capital, since capital has a zero profit when $0 < K < K^*$ – hence, one can expect no investment to occur. Thus, one could argue that those potential entrants whose borrowing constraint is lower than the minimum scale capital, $B^* \leq K^*$, will not enter the secondary sector, and will move to unemployment. The proportion of potential entrants for whom $B^* > K^*$, will be defined as θ.

Note that in the two-sector model of the previous section all those workers who were unable to find jobs in the primary sector were able to find a job in the secondary sector if they were willing to work for a wage equal to the marginal product of their labour. However, in the three-segment model of this section, barriers to entry into the secondary sector means that only a fraction, θ, of those who are unable to find jobs in the primary sector are able to enter the secondary sector. Therefore:

$$p_s = \theta(1 - p_p) \qquad (18)$$

That fraction, θ, is itself a function of the barriers of entry – the higher the barriers to entry, the lower the fraction. In terms of equations (16) and (17), the lower B^* is and the higher K^* is, the higher is the barrier to entry into the secondary sector and therefore the lower will θ be.

This implies that $(1 - p_p - p_s)$ is the proportion of positions that the primary and secondary sectors would have supplied, *had there not been barriers to entry in the secondary sector*. It

also means that, in this model, p_p and p_s are expressed as ratios of $F_p + F_s$, + U (which now comprises the labour force), with U being the involuntarily unemployed.

With the above, and similar to equation 2 above, the sum of the present value of expected primary, secondary and tertiary sector income in the economy is (where the zeros represent the zero wage earned by the unemployed):

$$PV = [(1 - q_p - d_1)w_p/r + (q_p + d_1)w_s/r]p_p + [(1 - q_s)w_s/r + q_s(0)/r]p_s + [(1 - q_u)(0)/r + q_u p_p w_p/r + q_u p_s w_s/r](1 - p_p - p_s) \quad (19)$$

In equilibrium, outflows from the primary sector need to equal inflows into the primary sector from the third segment (unemployed). Thus, $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$, which also means that $q_u = (q_p + d_1)/(1 - p_p - p_s)$.

In addition, the outflow from the secondary sector needs to equal inflow into the secondary sector from both the primary sector and the unemployed segment. Thus, $q_s p_s = (q_p + d_1)p_p + q_u p_s(1 - p_p - p_s)$, which (after reorganising) implies that $(q_p + d_1)p_p = q_s p_s - q_u p_s(1 - p_p - p_s)$ (which also equals $q_u p_p(1 - p_p - p_s)$ – see previous paragraph).

Assuming that the unemployed receive no income, it means that in this case too $\alpha/(d_2 - d_1) = (PV_p - PV_s)$ (compare equation 5). The present values of primary and secondary work are:

$$PV_p = (1 - q_p - d_1)p_p w_p/r + q_u p_p(1 - p_p - p_s)w_p/r \quad (20)$$

$$PV_s = ((q_p + d_1)p_p w_s/r + (1 - q_s)w_s/r + q_u p_s(1 - p_p - p_s)w_s/r \quad (21)$$

Therefore:

$$\alpha r/(d_2 - d_1) \leq (1 - q_p - d_1)p_p w_p + q_u p_p(1 - p_p - p_s)w_p - ((q_p + d_1)p_p w_s + (1 - q_s)w_s + q_u p_s(1 - p_p - p_s)w_s)$$

which reorganises as:

$$\alpha r/(d_2 - d_1) \leq ((1 - q_p - d_1)p_p + q_u p_p(1 - p_p - p_s))w_p - (((q_p + d_1)p_p + (1 - q_s) + q_u p_s(1 - p_p - p_s))w_s)$$

and after normalising on w_p yields:

$$w_p \geq \alpha r/(d_2 - d_1)((1 - q_p - d_1)p_p + q_u p_p(1 - p_p - p_s)) + (((q_p + d_1)p_p + (1 - q_s) + q_u p_s(1 - p_p - p_s))/((1 - q_p - d_1)p_p + q_u p_p(1 - p_p - p_s)))w_s \quad (22)$$

Using the equilibrium condition that $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$ (which also means $q_u = (q_p + d_1)/(1 - p_p - p_s)$), equation 22 simplifies to:

$$w_p \geq \alpha r/(d_2 - d_1)p_p + ((q_p + d_1)(p_p + p_s) + (1 - q_s))/p_p)w_s \qquad (23)$$

Now recall that $p_s = \theta(1 - p_p)$ and substitute it into equation 23 to yield:

$$w_p \geq \alpha r/(d_2 - d_1)p_p + (q_p + d_1)(1 - \theta)w_s + ((q_p + d_1)\theta + (1 - q_s))w_s/p_p \qquad (24)$$

Equation 24 represents the effort supply function in the three-segment model. As was the case with the two-sector model with no involuntary unemployment, an increase in p_p would cause w_p to decrease; and the slope of the effort supply function becomes flatter the larger p_p becomes.

Note that, unlike in equation 8, the quit rates do not disappear from equation 24. The reason for this is that the existence of barriers to entry into the secondary sector cause θ in equation 24 to be smaller than one (i.e. $\theta < 1$).[20]

6.3.2 The job-offer relationship and the price- and wage-setting relationships

Equations 9 to 11 from above remain unchanged, with equation 11' below subscripted for the primary sector:

$$w_p = ((\varepsilon - 1)/\varepsilon)(MPL) = ((\varepsilon - 1)/\varepsilon)(b(E_p)) \text{ with } b' < 0 \text{ and } w_p > 0 \qquad (9)$$

$$w_p = g(E_p) = \gamma(b(E_p)) \text{ with } g' < 0,\ \gamma = (\varepsilon - 1)/\varepsilon \text{ and } w_p > 0 \qquad (10)$$

$$p_p = (h/\gamma)w_p = (h/\gamma)g(E_p) \quad \text{or} \quad w_p = p_p\gamma/h \qquad (11')$$

Therefore, there is a positive relationship between p and w_p, but a negative relationship (given that $g < 0$) between E_p and p_p (as E_p increases, w_p decreases, causing p_p to also decrease). Equation 10 represents, again, the price-setting relationship, while equation 11' represents the job offer relationship.

---

[20] That the first term containing q_p would equal zero if $\theta = 1$, is straightforward to see. In the case of the second, recall that $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$, which means $(q_p + d_1) = q_u(1 - p_p - p_s)$, with $(q_p + d_1)$ appearing in the second term on the right-hand side of equation 24 that contain q_p. If $\theta = 1$ then $p_p + p_s = 1$, so that $q_u(1 - p_p - p_s) = 0$, which also means $(q_p + d_1) = 0$.) In the literature (cf. Campbell and Orszag 1998:121), higher levels of employment and wages are associated with a higher quit rate – higher employment levels imply a higher probability of finding a job again once the worker quits (more about this in section 4, which compares the two models).

Substituting equation 11' into equation 24 yields the detailed wage-setting equation:

$$w_p \geq \alpha r/((d_2 - d_1)(h/\gamma)g(E_p)) + (q_p + d_1)(1 - \theta)w_s + ((q_p + d_1)\theta + (1 - q_s))w_s/((h/\gamma)g(E_p)) \quad \text{with } g' < 0 \qquad (25)$$

As E_p increases (and given that $g' < 0$), w_p increases.

6.3.3 Model summary

The model can be summarised as follows.

First, in p-w_p space there are two relationships (the sign within [ ] below indicates the sign of the p-w_p relationship):

A job offer relationship:

$$p_p = (h/\gamma)w_p \quad \text{or} \quad w_p = p_p\gamma/h \qquad [+] \quad (11')$$

An effort supply function:

$$w_p \geq \alpha r/(d_2 - d_1)p_p + (q_p + d_1)(1 - \theta)w_s + ((q_p + d_1)\theta + (1 - q_s))w_s/p_p \qquad [-] \quad (24)$$

which is distinguished by the presence of θ (a function of barriers to entry, B) and quit rates

Secondly, in E_p-w_p space there are two relationships (with $g' < 0$) (the [ ] indicates the sign of the E_p-w_p relationship):

A price-setting relationship:

$$w_p = g(E_p) \qquad [-] \quad (10)$$

A wage-setting relationship:

$$w_p \geq \alpha r/((d_2 - d_1)(h/\gamma)g(E_p)) + (q_p + d_1)(1 - \theta)w_s + ((q_p + d_1)\theta + (1 - q_s))w_s/((h/\gamma)g(E_p)) \quad \text{with } g' < 0 \qquad [+] \quad (25)$$

which also is distinguished by the presence of θ and quit rates.

In a similar fashion as in the previous section, equations 11' and 24, and 10 and 25 can be used to calculate the equilibrium values for w_p, p_p and E_p:

$$w_p = (q_p + d_1)(1 - \theta)w_s + ((-(q_p + d_1)(1 - \theta)w_s)^2 + 4(\alpha r/(d_2 - d_1) + (q_p + d_1)\theta + (1 - q_s))w_s\gamma/h)^{0.5}/2 \qquad (26)$$

$$p_p = h(q_p + d_1)(1 - \theta)w_s/\gamma + ((-h(q_p + d_1)(1 - \theta)w_s/\gamma)^2 + 4(\alpha r/(d_2 - d_1) + (q_p + d_1)\theta + (1 - q_s))w_s h/\gamma)^{0.5}/2 \quad (27)$$

$$E_p = (h(q_p + d_1)(1 - \theta)w_s/\gamma + ((-h(q_p + d_1)(1 - \theta)w_s/\gamma)^2 + 4(\alpha r/(d_2 - d_1) + (q_p + d_1)\theta + (1 - q_s))w_s h/\gamma)^{0.5}/2)F \quad (28)$$

Note that, unlike their two-sector equivalents (equations 13-15), equations 26-28 contain θ (a function of barriers to entry, B) and the quit rates. The implications of these are discussed in the next section. Together with the effort supply function, the job offer relationship then determines the equilibrium number of positions in the primary sector. In addition, recalling that $p_s = \theta(1 - p_p)$, one can calculate the employment level in the secondary sector:

$$E_s = \theta(1 - p_p)F \quad (29)$$

In the three-segment model the unemployed are involuntarily unemployed. Those who end up in the third segment and who cannot re-enter either the primary or the secondary sectors due to the presence of barriers to entry into both the primary and secondary labour markets, find themselves involuntarily unemployed.

Using equations 28 and 29, one can calculate the total equilibrium employment level in the economy ($E_p + E_s$), which equals the equilibrium level of positions filled, $F_p + F_s$. Hence

$$U = F - (F_p + F_s) \quad (30)$$

is the number of involuntary unemployed.

6.4 A comparison of the two models

The two-sector, three-segment model shows how the two-sector model can be expanded from a model that merely explains the allocation of labour between the primary and secondary sectors, to a model that caters for the possibility of involuntary unemployment on the aggregate level. The key difference centres on the following. In the two-sector model, workers who quit/lose a job in one of the sectors, circulate back to a job in the other sector. In the three-segment model, workers who quit/lose a job in one of the two employing sectors do not necessarily find a job again and may end up being unemployed. Some workers might also never have worked (and remain unemployed).

The main reason why workers end up unemployed is the existence of barriers to entry such as a lack of physical and human capital discussed above. (If there are no barriers to entry into the secondary sector, the three-segment model reverts to the two-sector model.)

To compare the two models, consider equations 13-15 and 26-30. In the two-sector model quit rates do not play a role:

$$w_p = (\alpha r\gamma/(d_2 - d_1)h + w_s\gamma/h)^{0.5} \quad (13)$$

$$p = (\alpha rh/(d_2 - d_1)\gamma + w_s h/\gamma)^{0.5} \quad (14)$$

$$E_p = (\alpha rh/(d_2 - d_1)\gamma + w_s h/\gamma)^{0.5} F \quad (15)$$

In the three-segment model, barriers to entry as well as quit rates have an important role:

$$w_p = (q_p + d_1)(1 - \theta)w_s + ((-(q_p + d_1)(1 - \theta)w_s)^2 + 4(\alpha r/(d_2 - d_1) + (q_p + d_1)\theta + (1 - q_s))w_s\gamma/h)^{0.5}/2 \quad (26)$$

$$p_p = h(q_p + d_1)(1 - \theta)w_s/\gamma + ((-h(q_p + d_1)(1 - \theta)w_s/\gamma)^2 + 4(\alpha r/(d_2 - d_1) + (q_p + d_1)\theta + (1 - q_s))w_s h/\gamma)^{0.5}/2 \quad (27)$$

$$E_p = (h(q_p + d_1)(1 - \theta)w_s/\gamma + ((-h(q_p + d_1)(1 - \theta)w_s/\gamma)^2 + 4(\alpha r/(d_2 - d_1) + (q_p + d_1)\theta + (1 - q_s))w_s h/\gamma)^{0.5}/2)F \quad (28)$$

$$E_s = \theta(1 - p_p)F \quad (29)$$

$$U = F - (F_p + F_s) \quad (30)$$

Compared to the two-sector model, the presence of the quit rate q_p in the three-segment model's equations implies higher equilibrium values for w_p, p_p and E_p.[21]

In the literature (cf. Campbell and Orszag 1998:121), higher levels of employment and wages are associated with a higher quit rate – higher employment levels imply a higher probability of finding a job again once the worker quits. In two-sector model equilibrium, quit rates (as well as d_1, i.e. the probability of being laid-off for shirking while not actually shirking) do not affect w_p, p_p and E_p because in equilibrium the flow into the primary sector equals the flow out of the primary sector – those who quit find jobs in the secondary sector and are replaced, in turn, by workers moving from the secondary to the primary sector.

However, because of entry barriers in the secondary sector in the three-segment model, the flows into and from the primary sector are is not necessarily equal. This implies a relationship

[21] Why is this so? With $\gamma > h$ in all realistic scenarios, a higher q_p means that the third term on the right-hand side of equations 26-29 that contain q_p ($4(q_p + d_1)\theta w_s\gamma/h$) will always be larger than the second term that also contains q_p (for instance $-(q_p + d_1)(1 - \theta)w_s)^2$ in equation 26), leaving the net effect of these two terms as a positive value. With the first term on the right-hand side also containing q_p, the net effect of the three terms on the right-hand side containing q_p, will be positive, meaning higher equilibrium values for w_p, p_p and E_p. (The only exception to this scenario would be the primary sector goods market approximates an almost perfectly competitive market, contrary to the assumptions of this model.)

between quitting and w_p, p_p and E_p. In the three-segment model, barriers to entry mean that $\theta < 1$ (θ being a function of barriers to entry B). If $\theta = 1$, then all the terms containing q_p in equations 26-28 would disappear by virtue of being equal to zero,[22] which will also mean that q_p would have no effect. Thus, in this model the presence of barriers to entry (which cause $\theta < 1$) also ensure that q_p has an effect on w_p, p_s and E_p. Higher levels of employment in the primary sector imply that should a worker quit, the probability of ultimately finding a job again in the primary sector is higher, which, in turn, may engender a greater willingness on the part of primary sector workers to quit. Hence the positive relationship between quit rates and p_p and E_p.

Unlike the two-sector model where all workers are employed either in the primary or the secondary sector, in the three segment model $p_p + p_s \leq 1$ with $\theta < 1$. The higher the barriers to entry B, the lower p_p *and* p_s will be, hence (using equations 26, 27 and 28), the lower w_p and E_p will be. [23] Thus, barriers to entry mean fewer positions will be filled in both the primary and secondary sectors; employment will thus be lower. It also means wages in the primary sector will be lower than in the two-sector model.

Furthermore, note that the higher the quit rate q_s from the secondary sector, the lower are w_p, p_s and E_p. In the three-segment model, quitting from the secondary sector means that the worker moves towards unemployment, while in the two-sector model it meant that the worker circulates back to the primary sector. For given quit rates from the primary and tertiary sectors ('tertiary quitting' being quitting from unemployment and thus moving back to either primary or secondary sector employment), a higher quit rate in the secondary sector means a higher probability of ending up without a job, even if one starts out in the primary sector. Thus, a higher quit rate from the secondary sector depresses wages, employment and the number of jobs in the primary sector.

6.5 A graphical representation of the models

Figure 9 is a graphical presentation of the models discussed above. It shows employment in the two employing sectors on the horizontal axis and real wages W on the vertical axis. Primary sector employment is measured rightward from the vertical axis (marked W_P), while

[22] Why the first two terms containing q_p would equal zero if $\theta = 1$, is straightforward to see. In the case of the third, recall that $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$, which means $(q_p + d_1) = q_u(1 - p_p - p_s)$, with $(q_p + d_1)$ appearing in the third term on the right-hand side of equations 26-28 that contain q_p. If $\theta = 1$ then $p_p + p_s = 1$, so that $q_u(1 - p_p - p_s) = 0$, which also means $(q_p + d_1) = 0$.

[23] The logic is as follows: Higher barriers mean a lower θ, and the lower θ, the higher will be the first term on the right-hand side of equations 26-28 containing θ, but also the lower will be the second and third terms on the right-hand side of equations 26-28 containing θ. The effect of the second and third terms will exceed that of the first, which means that the net effect of these three terms on w_p, p_p and E_p in a case of a lower θ is negative. With both p_p and θ being lower, p_s will also be lower.

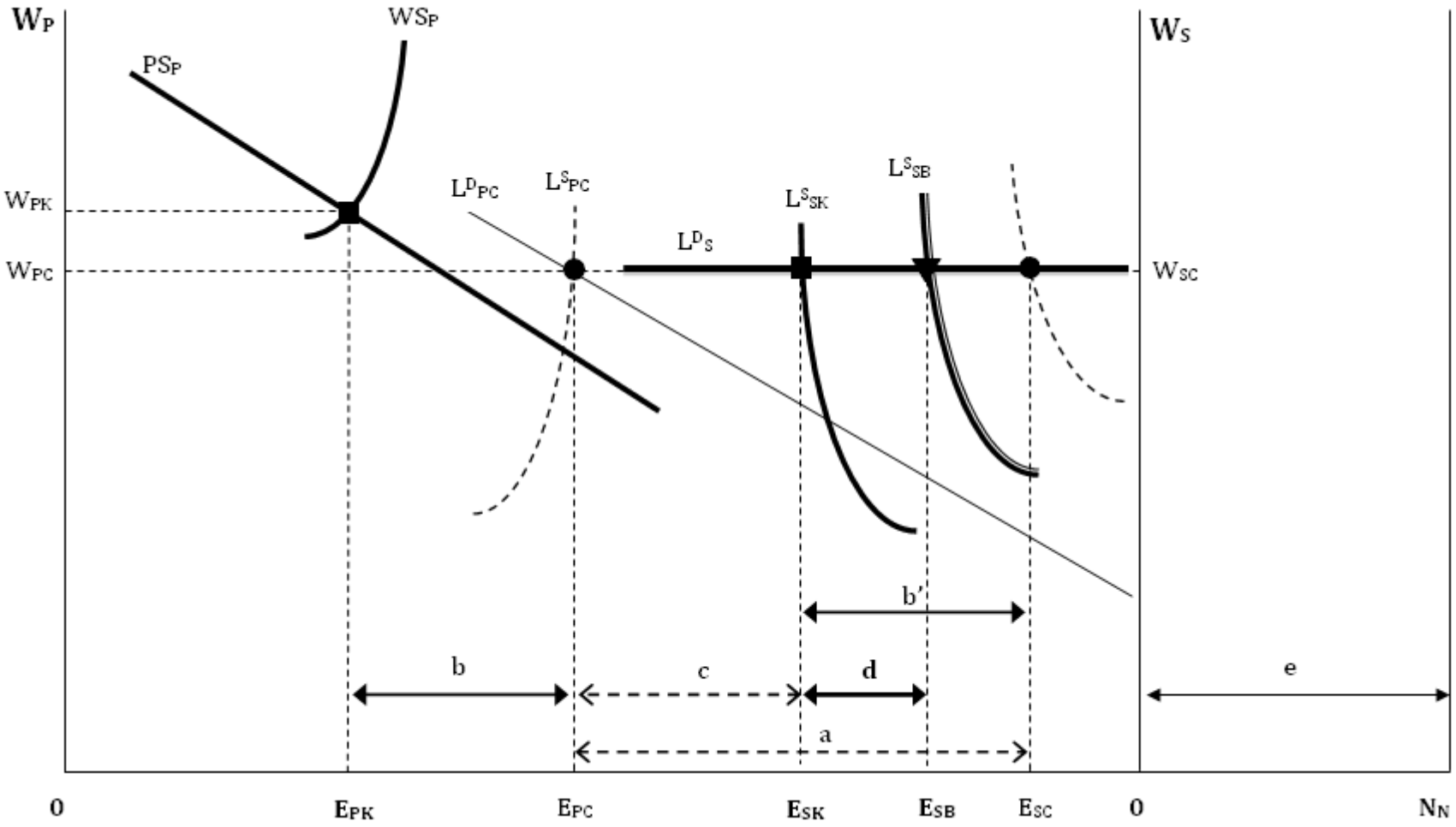
secondary sector employment is measured leftward from the vertical axis (W_S). N_N represents the working-age population. Distance *e* shows those who are not economically active.

Suppose, to start off, there is a perfectly competitive labour market with no market power and no efficiency wages. The wage paid in the primary and secondary sectors would be equal (i.e. there is no real distinction between the primary and secondary sectors). L^S_{PC} and L^D_{PC} represent labour supply and demand in a perfectly competitive (subscript C) labour market among firms in the primary sector, while L^S_{SC} and L^D_S represent labour supply and demand in the secondary sector. L^D_S is horizontal, following the simplifying assumption that the marginal product of labour in the secondary sector is constant.[24] Because the markets are perfectly competitive, wages in the primary and secondary sectors would be the same, $W_{PC} = W_{SC}$, with E_{PC} and E_{SC} being the corresponding employment levels in the primary and secondary sectors. The distance marked *a* represents those workers who would be voluntarily unemployed – they could always find work at the prevailing wage W_{SC} (i.e. if they are willing to reduce their reservation wages).

Now suppose the economy is Neo-Keynesian, with market power and efficiency wages in the primary sector. This produces the two-sector Neo-Keynesian model (subscript K), still with no barriers to entry into the secondary sector. The wage-setting (WP_P) and price-setting (PS_P) relationships in the primary sector will, due to effort behaviour, establish a wage W_{PK} that is higher than W_{PC}. Employment in the primary sector, at E_{PK}, will be lower compared to the perfectly competitive case, at E_{PC}. The difference in the number of workers being employed in the primary sector equals distance *b* in figure 9: $b = E_{PC} - E_{PK}$. Workers who are not accommodated in the primary sector, are diverted to and employed in the secondary sector. Thus, labour supply in the secondary sector is L^S_{SK} and *b*' (the horizontal leftward displacement from L^S_{SC} to L^S_{SK}) equals distance *b* (the quantity of workers relocated from the primary sector). Notice that in this model distance *a* equals distance $b + c$; since all these unemployed workers can find employment in the secondary sector at wage W_{SC} should they wish so (i.e. if they lower their reservation wage), they are voluntarily unemployed.

[24] Assuming a constant marginal product of labour for the secondary sector is not an altogether unrealistic assumption. Berry (2001:7) argues that large and medium enterprises usually have an amount of capital, which compliments a number of workers, in order to produce output. However, since a piece of capital has been designed for a specific (maximum) number of workers, increasing the number of workers, very quickly leads to a decrease in the marginal product of labour. However, by their very nature firms in the informal sector are very small, and the capital needed is replicable on a small scale (i.e. in the extreme case of one-person firms (own-employment) it is not the case that for instance a second workers is added to a given set of capital in a single small firm, but rather that the second workers can set up his or her own firm and replicate the capital – each worker is therefore the first worker and there is not really a second worker that can decrease the marginal product of labour. A similar point can be made for firms employing say two or three workers since with two or three workers, there is not much scope to decrease the marginal product of labour, particularly if the capital is replicable on a small scale. Berry (2001:7) argues that the flat marginal product of labour and thus the flat labour demand for informal sector workers has been well verified given the expandability of the informal sector. Of course, as the discussion below will indicate, there might be financial constraints on acquiring that minimal amount of capital, which might limit the size of the effective labour supply.

Figure 9 – Unemployment in the theoretical three-segment model



Next we introduce barriers to entry into the secondary sector (for simplicity we ignore barriers to entry into the primary sector). Given the nature of 'effort behaviour' in the primary sector, as before a quantity of workers equal to b will not be accommodated in the primary sector (compared to the perfectly competitive case). However, the presence of barriers to entry in the secondary sector means that labour supply in the secondary sector will be at L^S_{SB}. – lower than the previous case's L^S_{SK}. A quantity of workers equal to distance d will be involuntarily unemployed. This constitutes the third sector/segment in the model.

Unlike the case of the perfectly competitive market where workers can simply offer their labour at a lower wage, in a market with efficiency wages (with firms paying a wage to ensure effort), firms in the primary sector set both wages and prices. Hence, workers cannot increase employment in the primary sector by offering to work for a lower wage. In addition, even if unemployed workers are willing to work in the secondary sector for a wage equal to the marginal product of labour, barriers to entry prevent them from doing so.

The workers represented by distance c still are *voluntarily* unemployed. Even in the case of a perfectly competitive market, their reservation wage would have been above the market wage – they would have preferred unemployment even in the case of a perfectly competitive market. Note that the quantity of workers $b + d$ are willing to work in either the primary or the secondary sector at a wage of $W_{PC} = W_{SC}$, but are prevented from doing so due to the payment of efficiency wages in the primary sector and the existence of barriers of entry in the secondary sector respectively.

6.6 Conclusion and potential policy implications

To create a theoretical model that explains the dual nature of the South African labour market (with its formal and informal sectors) and the simultaneous existence, indeed persistence, of very high unemployment, this paper draws on the dual labour market model of Bulow and Summers (1986) and the suggestion by Kingdon and Knight (2004) as well as work by Grimm, Krüger and Lay (2011) that show that barriers to entry exist into the informal sector. Following the latter authors, such barriers are defined as the interaction of a borrowing constraint (itself the result of the asymmetric information faced by lenders in financial markets) and the minimum scale of capital needed to earn a high return.

In this way we develop a three-segment model comprising two sectors – a primary ('secure jobs') and a secondary ('less-secure jobs') sector/segment – as well as a third segment that comprises the unemployed.

The model shows:

1. How a primary sector characterised by efficiency wage and labour union behaviour as well as a mark-up due to high transport cost, can explain the dual nature of the labour market.
2. How barriers to entry faced by potential entrants into the secondary sector can prevent workers from entering the secondary sector. This constrains the effective supply of labour to the secondary sector.
3. How, as a result, these workers end up being (involuntarily) unemployed in long-term macroeconomic equilibrium. The secondary sector does not simply absorb all those who cannot find employment in the primary sector.
4. Disturbances and fluctuations in the primary sector, for example, would spill over into the secondary sector *and* the third segment (comprising the unemployed).

From a policy point of view the above suggests that there is no single or 'silver bullet' solution to address the dual nature of the labour market or the unemployment problem. The solution is not as easy as, for instance, simply decreasing wage levels to render labour cheaper (a solution often proposed in some academic and corporate circles).

More specifically with respect to the secondary sector, the analysis shows that it cannot merely adjust wages to soak up all of the unemployed looking for employment. In addition, one can also not just expect that all those wanting to be self-employed can in fact do so – there might be barriers preventing them from doing so. Indeed, if the assumptions on which the above model draws hold in the South African reality, then a solution to the unemployment problem will require a multipronged approach that need to involve policies addressing product and labour market structures and behaviour in the primary sector, as well as policies addressing the numerous barriers to entry, such as borrowing constraints, that potential entrants into the secondary sector face.

7. Final thoughts: Towards a macroeconomic research agenda

The existence of labour market segmentation is a well-known finding in labour market research, and poverty research frequently highlights labour market marginalisation. Given the interest in the unemployment-poverty nexus, interest has focused on segmentation between the formal sector and informal sector/economy. Therefore, in referring to Layard-type models in the SA context, the primary and secondary sectors routinely are taken to be the formal and informal sectors respectively (e.g. Kingdon and Knight 2004).

This paper has argued for a macroeconomic framework that incorporates the diversity of economic activities ranging from the formal sector to the informal sector and thereby provides a suitable basis for macroeconomic policy in the peculiar South African context. It illustrated

such a framework by deriving a novel three-segment model that explicitly incorporates a primary sector (formal sector), the secondary sector (informal economy) and a tertiary (unemployed-persons) segment, as well as labour market entry and mobility barriers.

With respect to long-run labour market equilibrium, the model explains the existence of involuntary unemployment and a tertiary segment in such an equilibrium. The existence of labour-market entry and mobility barriers are central to this result, in addition to assumptions on imperfectly competitive product and labour markets (in New Keynesian fashion).

With respect to short-term shocks and dynamics, the three-segment model enables analyses of the impact of cyclical aggregate demand disturbances or aggregate supply shocks on employment conditions in *both* sectors – *and* on the tertiary, last-resort segment. This provides a systematic framework for situating inter-segment transitions and labour-market flows in a segmented macroeconomic model. Phenomena like hysteresis can also be introduced.

The model above paves the way for an unemployment-oriented macroeconomic research and policy agenda that focuses on some of the following aspects:

1. The differential impact of supply and demand shocks (macroeconomic cycles) on employment and unemployment in the primary (formal) and secondary (informal) sectors, and
2. How the nature and severity of structural conditions such as segmentation and various entry barriers can soften or aggravate this impact.
3. The impact of macroeconomic growth or stagnation on employment (and unemployment) in both the primary (formal) and secondary (informal) sectors.
4. The impact of labour market structural conditions (segmentation; entry barriers, wage differentials) on the employment effect of macroeconomic growth.
5. The extent to which labour market structural conditions are a constraint on macroeconomic growth.

References

Agénor, PR and Montiel, P. *Development macroeconomics.* 2nd ed. Princeton (New Jersey): Princeton University Press.

Berry, A. 2001. The role of the small and medium enterprise sector in Latin America: Implications for South Africa. TIPS Working Paper 5. Online: http://www.tips.org.za/files/421.pdf

Blanchard, O. 2005. *European unemployment: The evolution of facts and ideas.* Working Paper 11750, National Bureau of Economic Research, Cambridge, Mass.

Bulkley, G and Myles, GD. 1996. Trade unions, efficiency wages and shirking. *Oxford Economic Papers,* 48(1996): 75-88.

Bulow, JI and Summers, LH. 1986. A theory of dual labor markets with application to industrial policy, discrimination and Keynesian unemployment. *Journal of Labor Economics*, 4(3)(Part 1): 376-414.

Cahuc, P and Zylberberg, A. 2004. *Labor Economics*. MIT Press.

Campbell, C and Orszag, JM. 1998. A model of the wage curve. *Economics Letters*, 59 (1998): 119–125.

Carlin, W and Soskice, D.2005. *Advanced Macroeconomics*. Oxford University Press.

CDE (Centre for Development and Enterprise). 2013. Graduate unemployment in South Africa: A much exaggerated problem. *CDE Insight*. April 2013. http://www.cde.org.za/publications/jobs-growth/83-jobs-and-growth/403-graduate-unemployment-in-south-africa-a-much-exaggerated-problem

Clower, R.W. 1965. The Keynesian Counter Revolution: a theoretical appraisal. in R.W. Clower (ed.) *Monetary Theory.* Middlesex: Penguin.

Fedderke, J and Naumann, D. 2011. An analysis of industry concentration in South African manufacturing, 1972–2001. Applied Economics, 43(22), Special Issue: The Applied Economics of Industry, 2919-39.

Fourie, FCvN. 2011. *The South African unemployment debate: Three worlds, three discourses?* Working Paper 63, SALDRU, University of Cape Town.
Or: Working Paper 1, REDI3x3. Available at: http://www.redi3x3.org

Fourie, FCvN and Burger, P. 2015. *How to think and reason in macroeconomics.* 4th edition. Cape Town: Juta.

Grimm, M, Krüger, J and Lay, J. 2011. Barriers to entry and returns to capital in informal activities: Evidence from Sub-Saharan Africa. *Review of Income and Wealth,* Series 57, Special Issue, May 2011: S27-53.

Grimm, M, Van der Hoeven, R and Lay, J. 2011. *Unlocking potential: Tackling economic, institutional and social constraints of informal entrepreneurship in Sub-Saharan Africa: Main findings and policy conclusions.* International Institute of Social Studies, Erasmus University.

Kingdon, GG and Knight, JB. 2004. Unemployment in South Africa: the nature of the beast. *World Development,* 32(3): 391-408.

Knell, M. 2014. Efficiency wages, staggered wages, and union wage-setting. Oxford Economic Papers Advance Access published March 24, 2014.

Layard, R., Nickell, S and Jackman, R. 1991 and 2005. *Unemployment: macroeconomic performance and the labour market*. Oxford: Oxford University Press.

Leijonhufvud, A. 1967. Keynes and the Keynesians – A suggested interpretation. *American Economic Review*, 57(2): 401-10.

Pereau, J-C and Sanz, N. 2006. Trade unions, efficiency wages and employment. *Economics Bulletin,* 10(4): 1-8.

Posel, D, Casale, D and Vermaak, C. 2014. Job search and the measurement of unemployment in South Africa. *South African Journal of Economics*, 82(1): 66-80.

Romer, D. 2012. *Advanced Macroeconomics*. 4th ed. McGraw-Hill Irwin. New York.

Statistics South Africa. 2009. Labour Force Survey - Historical Revision: September Series 2000 to 2007. Statistical release P0210.

Statistics South Africa. 2014. Revised QLFS trends 2008-2013 corrected. Excel datasheet available at www.statsSA.gov.za

Summers, LH. 1988. Relative wages, efficiency wages, and Keynesian unemployment. *American Economic Review,* 78(2): 383-8.

Van der Berg, S and Van Broekhuizen, H. 2012. *Graduate unemployment in South Africa: A much exaggerated problem.* Working Paper 22/2012, University of Stellenbosch. http://resep.sun.ac.za/index.php/research-outputs/stellenbosch-working-papers/wp2012/

The **Research Project on Employment, Income Distribution and Inclusive Growth (REDI3x3)** is a multi-year collaborative national research initiative. The project seeks to address South Africa's unemployment, inequality and poverty challenges.

It is aimed at deepening understanding of the dynamics of employment, incomes and economic growth trends, in particular by focusing on the interconnections between these three areas.

The project is designed to promote dialogue across disciplines and paradigms and to forge a stronger engagement between research and policy making. By generating an independent, rich and nuanced knowledge base and expert network, it intends to contribute to integrated and consistent policies and development strategies that will address these three critical problem areas effectively.

Collaboration with researchers at universities and research entities and fostering engagement between researchers and policymakers are key objectives of the initiative.

The project is based at SALDRU at the University of Cape Town and supported by the National Treasury.

Consult the website for information on research grants and scholarships.

Tel: (021) 650-5715

www.REDI3x3.org

REDI3X3

Research Project on Employment, Income Distribution & Inclusive Growth

Towards an integrated response to unemployment, inequality and poverty